8-1995

Autocorrelation in Single-Subject Data: A Meta-Analytic View

Laura L. Methot
Western Michigan University

AUTOCORRELATION IN SINGLE-SUBJECT DATA:
A META-ANALYTIC VIEW

by

Laura L. Methot

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
August 1995

AUTOCORRELATION IN SINGLE-SUBJECT DATA: A META-ANALYTIC VIEW

Laura L. Methot, Ph.D.

Western Michigan University, 1995

Recent work by Huitema and McKean (1991, 1994a, 1994b, 1994c, in press) has shown that the most frequently used statistical methods for performing conventional time-series analyses lead to gross distortions of results when these approaches are applied in the context of the typical behavioral research study. Most of these problems could be avoided if researchers were aware that the time-series methods recommended in many areas are not generally needed. The appropriate evidence regarding the need for complex time-series methods requires a meta-analysis of the autocorrelation present in behavioral studies. The project involved: (a) sampling several hundred research articles published in the *Journal of Applied Behavior Analysis* (*JABA*) during 1990-1995, (b) extracting the quantitative data from the time-series data displays in 200 sample articles, (c) analyzing each data set using recently developed statistical methods, and (d) performing a meta-analysis of the individual analyses using a new methodology developed for time-series data.

Conventional lag-1 autocorrelation estimates r_1 (known to be negatively biased) were computed on (a) the raw data and (b) the residuals obtained from regressing the raw data on time in order to remove the linear trend. New unbiased autocorrelation estimators (r_{F1} and r_{HM}) were applied to both the raw and detrended data. New tests of the homogeneity of the coefficients in the distributions of r_{HM} and r_{F1} were computed, and confidence intervals on the means of the distributions were computed.

Finally, new tests of the significance of the obtained r_{F1} and r_{HM} values were computed.

Results indicate that recent conclusions regarding the presence of the autocorrelation "problem" in typical behavioral experiments should be questioned. When the unbiased estimator (r_{F1}) is used, there appears to be a significant proportion of behavioral data sets exhibiting autocorrelation, generally in the positive direction. Most of the apparent positive autocorrelation can be explained by linear trend in the data. When the data are appropriately detrended the autocorrelation distribution is similar to that expected by chance when sampling from a population of independent (nonautocorrelated) errors. The conclusion that behavioral data are not highly autocorrelated is drawn. In most cases the explanation for "apparent" autocorrelation is not that the errors are autocorrelated. Rather, the statistical model used to explain the data has been misspecified. Artificial autocorrelation is to be expected if trend (linear or otherwise) and other deterministic components in the data have not been specified in the statistical model. Conventional general linear model solutions are satisfactory for many studies in applied behavior analysis.

UMI Number: 9600945

UMI

300 North Zeeb Road
Ann Arbor, MI 48103

ACKNOWLEDGMENTS

Many people are deserving of recognition for their contributions to the completion of this dissertation. First and foremost, to my advisor Dr. Bradley Huitema, I extend my sincere appreciation for his constant support, guidance, and timely feedback. Without the exemplary mentoring of Dr. Huitema this project would yet to be off the ground.

To my committee members, Drs. Alyce Dickinson, Alan Poling, and Joseph McKean, I extend appreciation for their versatility in timing and their acceptance of the topic of this study. Thanks go also to Drs. Jack Michael and Richard Spates for their participation in my oral examination and for their helpful comments on the final manuscript.

Finally, I wish to thank my family members who modeled persistence every step of the way. I dedicate this work to Pete Maruh, who provided an always safe haven from the daily demands of academe.

Laura L. Methot

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Figures--Continued

INTRODUCTION

The appropriate analysis of single-subject data has been debated in numerous articles throughout the history of behavior analysis (e.g., Gentile, Roden & Klein, 1962; Jones, Vaught & Weinrott, 1977; Matyas & Greenwood, 1991; Michael, 1974). On one side of the debate are the proponents of visual analysis of graphically presented data (Baer, 1977; Michael, 1974; Baer & Parsonson, 1978, 1992; Parsonson & Baer, 1986; Sidman, 1960), and on the other side are those who assert that more formal statistical analyses are necessary or useful aids in the analysis of behavioral data (Busk & Marascuilo, 1988; Gottman & Glass, 1978; Sharpley & Alavosius, 1988; Suen, 1987; Suen & Ary, 1987). Those in favor of more formal statistical analyses are further divided on the issue of the most appropriate formal methods. A brief overview of the history of these issues is provided in subsequent sections to create the relevant context for the problem investigated in this dissertation.

Analysis of Single-subject Data

Visual Analysis

The results of single-subject research are typically presented in graphic form for visual inspection. Data are plotted over time, baseline and intervention phases are clearly identified, and the goal is to determine whether changes in experimental conditions are associated with changes in the behavior from one phase to another. The merits of visual analysis are extolled by Kazdin in the following passage:

> Visual inspection is regarded as a relatively *un*refined and *in*sensitive criterion for deciding whether the intervention has produced a reliable change. The unsophisticated features of the method are regarded as a

> virtue. Because the criterion is somewhat crude, only those interventions that produce very marked effects will lead the scientific community to agree that the intervention produced a change. Weak results will not be regarded as meeting the stringent criteria of visual inspection. Hence, visual inspection will serve as a filter or screening device to allow only clear and potent interventions to be interpreted as producing reliable effects (1982, p. 232).

This passage suggests that visual inspection is viewed as superior to a more formal statistical analysis because the latter may result in the identification of small or weak effects. In fact, behavioral researchers assert that one undesirable characteristic of statistical analysis is its sensitivity in detecting weak intervention effects (e.g., Baer, 1977; Johnston & Pennypacker, 1991). Human judgement of intervention effects via visual analysis of graphed data is not, however, flawless. The lack of concrete decision rules for determining intervention effects has been cited as a potential weakness of the visual inspection method (Kazdin, 1982, p. 239). Further, the reliability of conclusions across judges appears moderate at best (DeProspero & Cohen, 1979; Jones, Weinrott, & Vaught, 1978). Finally, some methodologists suggest that trends and variability in the data may mask effects and render them undetectable by means of visual analysis (e.g., Kazdin, 1982, p. 240). It is not uncommon to encounter phases in which performance gradually deteriorates or improves over time. Some methodologists argue that determining whether trend within one phase is simply a continuation of trend from a previous phase, or if an actual treatment effect is present, can be difficult to judge visually. Unless there is an obvious change in level or slope from one phase to the next, treatment effects might go undetected (Barlow & Hersen, 1984). While most behavior analysts argue that the answers to these problems lie in generating better experimental control to minimize intrasubject response variability and to maximize intraphase response stability (e.g., Johnson & Pennypacker, 1991; Michael, 1974; Sidman, 1960), others suggest that this is not always possible, particularly when operating within the constraints of applied research (e.g., Barlow & Hersen, 1984).

Statistical Analysis

Statistical analyses are used to describe and summarize data and to help determine whether different experimental conditions are associated with differences in the nature of the behavior under two or more levels of an independent variable. Statistical procedures have generally been employed with sets of data collected from groups of subjects, but may also be used to compare data gathered from a single subject measured repeatedly across time. Data collected in this manner from a single subject are frequently described as a time-series and the statistical analysis of such data is often described as a time-series analysis.

Rationale for Use of Formal Statistical Analyses With Applied Behavioral Data

The rationale for supplementing visual with statistical analysis in behavioral studies generally follows one of two lines of reasoning. First, detecting intervention effects through visual inspection of data requires a stable baseline. When baseline data show trends, the argument goes, intervention effects may be masked. Similarly, when nontrending baseline data show large variability it may be difficult to detect relevant functional relations (Kazdin, 1982, p. 242-244). The resultant argument is that statistical decision aids may be valuable in such cases to conclude that the intervention did or did not affect the behavior of interest.

A second rationale for supplementing a visual analysis with a statistical analysis is that small effects may be important (Kazdin, 1982, p. 244); it is an advantage rather than a disadvantage if statistical analyses are more sensitive than visual analysis for detecting treatment effects. It should be pointed out, however, that the data supporting the claim of higher sensitivity for statistical analysis in the case of behavioral data is not particularly convincing (Huitema, 1986b).

Statistical analyses for single-subject data have received much attention since Gentile et al. (1962) proposed that the analysis of variance (ANOVA) be employed in analyzing data from single-subject designs. This method of analysis entails comparison of the means of two or more phases based on the data collected from a single subject. Many other statistical methods have been proposed to compare behavior under different conditions. A general inferential question posed by all methods is, "Do the data provide enough evidence to conclude that the intervention had an effect?" The adequacy of the inferential statements associated with all statistical methods depends upon the extent to which the data conform to the assumptions of the statistical model underlying the analysis.

Assumptions Underlying Conventional Linear Models

Inferential statistical procedures involve fitting a model to the data. A statistical model is an equation that explains behavior on the dependent variable using one or more parameters. The models sparking the debate over the applicability of inferential procedures in the analysis of behavioral data are analysis of variance models.

There are three assumptions underlying the analysis of variance. It is assumed that the errors of the model are independent, that they have the same variance under each treatment condition, and that they are normally distributed. Central to the current debate regarding the analysis of behavioral data is the assumption of independence. In order to understand the assumption of independence it is helpful to look at the components of the ANOVA model. The ANOVA model in the case of a one-factor between subjects design is as follows:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

where Y_{ij} is the observed score for the *i*th subject in the *j*th group, μ is the population grand mean for all groups in the study, α_j is the effect of falling in the *j*th group, and ε_{ij} (called the error component) is the discrepancy between the *i*th subject's observed score and the mean (μ_j) for the *j*th group.

For any collection of sample raw scores, there will be a corresponding collection of residuals, defined as

$$Y_{ij} - \overline{Y}_j = e_{ij}$$

where Y_{ij} is the observed score for the *i*th subject of the *j*th group, $\overline{Y}_j$ is the mean of the group from which the score was drawn, and e_{ij} is the resulting residual, which may be plotted as a frequency distribution in the same manner as the raw scores. These residuals are estimates of the errors in the ANOVA model; they provide information on the extent to which the observed sample data conform to the three assumptions. The errors of the ANOVA model represent random variation that is unexplained by the other terms in the model.

Similarly, the simple linear regression model is written as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where Y_i is the observed score for the *i*th subject. β_0, the population intercept, is the elevation on Y when X is equal to zero. β_1, the population slope, is the number of units increase or decrease on Y for one unit increase on X. X_i is the level on the predictor variable associated with Y_i, and ε_i (the error component) is the discrepancy between the *i*th subject's observed score and the expected value of Y_i given X_i (i.e., $\beta_0 + \beta_1 X_i$). As with the ANOVA model, for any collection of sample raw scores there will be a corresponding collection of residuals associated with the regression model (defined as (Y_i - $\hat{Y}_i$ = e_i; where $\hat{Y}_i$ is the value of Y_i predicted from the fitted regression line). The residuals of the regression analysis are used to evaluate the adequacy of the regression model as a description of the data in the same way that the residuals of an ANOVA are used to evaluate the adequacy of the ANOVA model in describing data.

While the analysis of variance and regression models specified above were described as appropriate for explaining the behavior of various subjects in "group" designs, these models can be written in a modified form to explain the behavior of a single subject in a time-series design. In the case of the ANOVA model the behavior to be explained is Y_{tj}; this is the dependent variable score for the one subject in the study measured at time t under condition j. The explanatory terms are μ, α_j, and ε_{tj} where μ is the overall level of the time-series, α_j is the effect of the condition applied during phase j of the experiment, and ε_{tj} is the error defined as Y_{tj} - [$\mu + \alpha_j$]. It is assumed once again that the errors are independent with constant variance across all phases, and normally distributed. Table 1 presents an example of a set of residuals computed by subtracting the phase mean from each of a set of raw scores. The original raw scores and their corresponding residual values are plotted in Figures 1 and 2. As can be seen in Figure 1, the raw scores vary around the phase means of 9 and 59 while Figure 2 shows that the residual values vary around a mean of zero for each phase.

Correspondingly, the simple linear regression time-series model explains the behavior (Y_t) of the one subject in the study at time t rather than the behavior of various subjects at one point in time. The explanatory variable is time rather than X. Hence the model is written: $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$. It is assumed that the errors are independent with the same variance at each time point and normally distributed. Figure 3 shows a sample data set of 10 observations. An upward trend is clearly evident in this graph. The

Table 1

Example of Residuals Based on the ANOVA Model.

Observation number	Raw score Y_{ij}	-	Phase Mean $\overline{Y}_j$	=	Residual e_{ij}
1	11		9		+2
2	8		9		-1
3	7		9		-2
4	10		9		+1
5	9		9		0
6	80		59		+1
7	57		59		-2
8	58		59		-1
9	61		59		+2
10	59		59		0

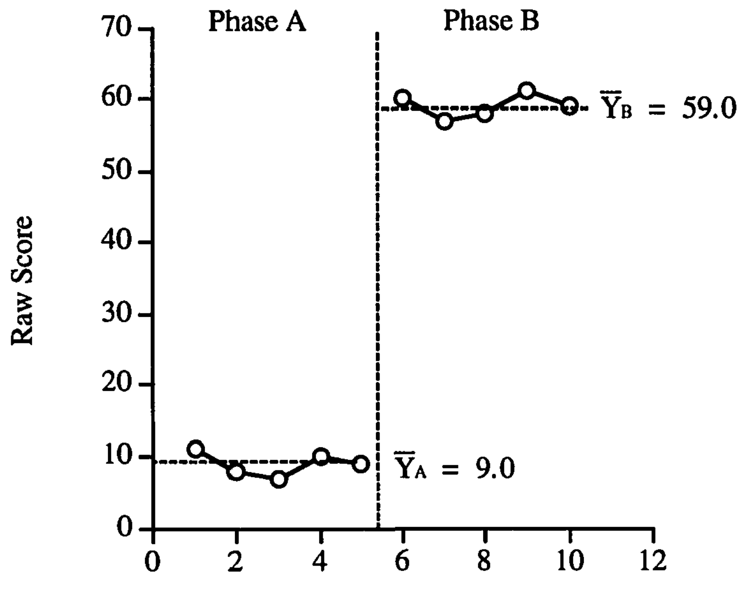

Figure 1. Example of Raw Scores Varying Around Phase Means for a Two-Phase Study.

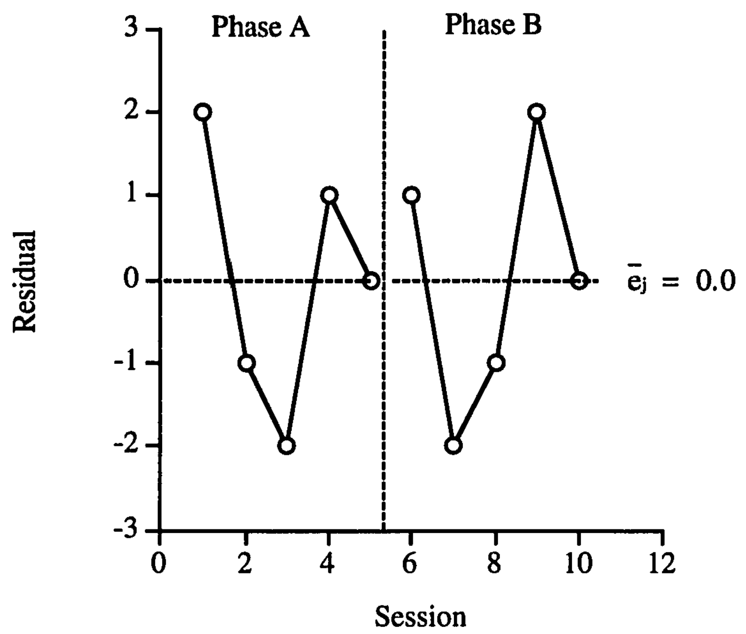

Figure 2. Example of Residual Values Varying Around Means of Zero for a Two-Phase Study.

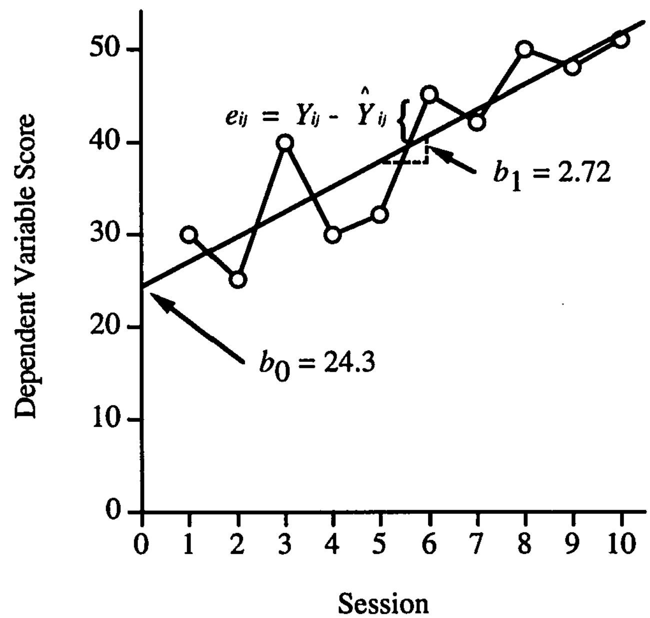

Figure 3. Example of Upward Trending Data With a Fitted Regression Line.

fitted regression line has a slope (b_1) of 2.72; that is, $\hat{Y}$ increases 2.72 units for every one unit increase on X. The sample intercept (b_0) is equal to 24.3, and the residual value is computed by subtracting $\hat{Y}$ (the level on Y predicted by the regression line) from the raw score Y. The terms b_0, b_1, and e_{ij} are sample estimates of the unknown population values β_0, β_1, and ε_i.

Violation of the independence assumption results in invalidation of the ANOVA and linear regression *F* tests because the probability value associated with the test will be conservative or liberal, depending on the degree and direction of dependency among the errors (Huitema, 1985). When applying inferential procedures such as ANOVA with between-group studies involving independently treated subjects, there is little question regarding the independence of the errors. The question arising with respect to behavioral time-series data, however, is "to what extent can the errors of data gathered over time from a single subject be considered independent?" This question can be investigated empirically by computing coefficients that estimate dependency among the errors of the relevant model. Such coefficients are called autocorrelations. If the errors are autocorrelated an alternative model should be considered. But it is important to thoroughly understand the nature of the "relevant" autocorrelation before considering the issues in selecting an alternative model.

The Autocorrelation Debate

A series of papers appearing in the 1970s and 80s argued for alternatives to conventional parametric statistical tests in the analysis of behavioral data (Busk & Marascuilo, 1988; Hartmann, Gottman, Jones, Gardner, Kazdin, & Vaught, 1980; Jones, Vaught, & Reid, 1975; Jones et al., 1977; Jones et al., 1978; Sharpley & Alavosius, 1988; Suen, 1987; Suen & Ary, 1987). The justification for this argument rested on the notion that behavioral data necessarily violate the independence assumption underlying the analysis of variance and other variants of the general linear model. The lack of independence was ostensibly illustrated empirically through the computation of non-zero autocorrelation coefficients on single-subject data sets (Busk & Marascuilo, 1988; Jones et al., 1977; Sharpley & Alavosius, 1988; Suen, 1987; Suen & Ary, 1987). Additionally, the logical stance that human performance at one point in time is influenced by performance at an earlier point was taken as evidence that behavioral data violate the independence assumption (Busk & Marascuilo, 1988; Jones et al., 1977; Sharpley, 1988).

Interrupted time-series models that contain parameters to explain the assumed nonindependence have been widely recommended as necessary alternatives for the statistical analysis of behavioral data (e.g., Jones et al., 1975, 1977; Hartmann et al., 1980). Huitema (1985, 1986a, 1986b, 1988) asserts that the recommendation to routinely use these techniques is unfounded. He notes that "because applied behavior data are collected across time it has been presumed that time-series analysis is required" (1986b, p. 225). A fundamental misunderstanding of the assumption of independence is the foundation for the continued erroneous arguments against the use of conventional general linear model procedures (including time-series regression intervention models and the analysis of variance) and in favor of formal (e.g., ARIMA) time-series analysis of behavioral data. Clearly earlier attempts to explicate this critical assumption (Huitema 1985, 1986a, 1988) have not been entirely successful. The purposes of the three following sections are to summarize the "autocorrelation problem" and its relationship to the independence assumption, review the body of literature that has been generated by the debate over this alleged problem, and describe the intent of the research carried out in this dissertation.

The Independence Assumption and Autocorrelation

It has been pointed out in all papers that argue for the use of complex time-series methods that the independence assumption will be violated in the case of behavioral time-series data. This presumption has been frequently supported with empirical evidence regarding the autocorrelation coefficients associated with behavioral data. Unfortunately, conceptual and computational errors appear to surround all of these empirical studies (Huitema, 1986a, 1988, 1990). The conceptual and computational issues associated with autocorrelation coefficients and independence are presented next.

The Meaning of Autocorrelation

Inaccurate verbal definitions of autocorrelation as well as computational inaccuracies characterize discussions of autocorrelation in the behavioral literature. Autocorrelation, or serial dependence, in a time series has been defined as follows: the level obtained at one point in the series is to some extent predictable based on knowledge of the levels at previous points in the series (Glass, Willson, & Gottman, 1975). But this definition is misleading; it has been pointed out (Huitema, 1985, 1986a) that the level of behavior at one point in the series is somewhat predictable from the previous data in the series even when there is absolutely no autocorrelation. It may be appropriate to avoid verbal definitions of autocorrelation and to instead focus on the nature of the computations involved in computing autocorrelation coefficients.

An autocorrelation coefficient is a correlation computed between the observations (or residuals) of a time series and the lagged values of that series (Huitema, 1994). It is possible to compute autocorrelations on raw data or the residuals of any number of statistical models, the relevant computation being dictated by the nature of the data and the question posed by the investigator.

Several papers (Huitema, 1985, 1986a, 1988) have pointed out that earlier investigators (e.g., Busk & Marascuilo, 1988; Jones et al., 1975; Jones et al., 1977; Jones et al., 1978) based their recommendations for using complex time-series analysis techniques on the faulty assumption that it is autocorrelation of the observed data that invalidates the use of procedures based on the general linear model, such as the analysis of variance. To state, though, that behavioral data are not necessarily plagued with an autocorrelation problem, as Huitema has done (1985, 1986a, 1988), has generated the erroneous logical position that (a) if there is no autocorrelation in behavioral data, then (b) people behave randomly over time, thus (c) the prediction and control of behavior is nonexistent (see Huitema, 1986b). He points out that this argument is faulty because zero autocorrelation of the errors ε_t of a statistical model does not imply that behavior is unpredictable. Behavior at any point in time in a series may be predicted using the mean of the data in a non-trending series, or by using a regression equation fitted to the data when a trend is present.

The presence of autocorrelation in the residuals of either ANOVA or simple linear regression generally means that the correct statistical model has not been identified. A regression model, for example, will provide a better fit to data trending over time than a model based on the mean. Similarly, some environmental determinant of the behavior under study can cause autocorrelation of the residuals if the environmental event is not included in the model. Thus, the logic that no autocorrelation means that behavior is random and unpredictable does not hold because the statistical models used to explain behavior contain both deterministic and random components. The lack of autocorrelation of the errors simply means that the error at one point in time is not predictable from previous errors; this says nothing about the predictive usefulness of the other terms in the model (Huitema, 1986a).

Computational Basis of Autocorrelation

Serial dependence of the errors of a model is assessed by computing a correlation between the residuals of the time series and the lagged values of the series. A lag-1 residual autocorrelation, for example, is derived by pairing residuals at time t with those at time $t+1$ and computing a correlation coefficient similar to (but not identical with) the Pearson coefficient. The result, which is labeled r_1, is a summary measure of the degree of linear dependence between contiguous values. Table 2 (adapted from Huitema, 1994) contains columns of residuals that are relevant in computing r_1 through r_{50}. That is, the values in column e_t are correlated with those in column e_{t+1} to compute r_1; those in column e_t are correlated with those in e_{t+2} to compute r_2, and so on. Huitema states that "it is possible to compute as many autocorrelations as lags, but generally only about one-sixth to one-fourth of all possible autocorrelations are computed" (1994, p. 430). This limit on the number of coefficients is recommended because the autocorrelation estimates become less stable as the lag increases. Note that the lag-1 autocorrelation in Table 2 is based on 199 pairs of numbers, but the lag-50 autocorrelation is based on 150 pairs of numbers (i.e., 200 observations in the original series minus 50 lags). One pair of numbers in the computation is lost for every lag, a fact which severely impinges upon the stability of sample autocorrelations for small data sets. The whole set of sample autocorrelation coefficients (e.g., r_1 through $r_{N/4}$, where N/4 is the number of lags considered in the computations) is sometimes called the sample autocorrelation function. The conventional formula for the lag-k autocorrelation coefficient is:

$$r_k = \frac{\sum_{t=1}^{n-k}\left(Y_t - \overline{Y}\right)\left(Y_{t+k} - \overline{Y}\right)}{\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)^2}$$

Tests of significance can be computed on the sample autocorrelation coefficients to determine whether the underlying population series (called the *process*) of errors of some specified model is autocorrelated. Many tests are available for this purpose (e.g., Bartlett, 1946; Box & Pierce, 1970; Huitema & McKean, 1991, 1994c, submitted; Ljung & Box, 1978). If the tests suggest that autocorrelation is present it is concluded that the assumption of independent errors is violated. The identification of autocorrelated errors means that the model of behavior (Y_t) is misspecified and that an alternative model is required to explain the data.

Table 2

Example of Basic and Lagged Residuals

t Time	e_t residual at Time t	e_{t+1} residual at Time $t+1$	e_{t+2} residual at Time $t+2$	...	e_{t+50} residual at Time $t+50$
1	e_1=-2.33	e_2= 4.67	e_3= 0.67		•
2	e_2= 4.67	e_3= 0.67	e_4=-5.33		•
3	e_3= 0.67	e_4=-5.33	e_5= 6.67		•
4	e_4=-5.33	e_5= 6.67	•		e_{200}=-4.3
5	e_5= 6.67	•	•		
•	•	•	•		
•	•	•	e_{200}=-4.3		
•	•	e_{200}=-4.3			
200	e_{200}=-4.3				

Alternative Intervention Models

If a researcher is interested in statistically analyzing the data from a behavioral intervention study, there are several approaches that can be taken. Among them are the ANOVA model, an interrupted time-series regression model, an ARIMA intervention model, and a new "double bootstrap" method. Each of these methods is most appropriate under different conditions. A brief description follows.

ANOVA Time-Series Model

If there is no apparent trend in the data, that is, the data appear stationary around some mean value within each phase of the study, then the ANOVA model may provide an appropriate fit. This model has been found to be adequate for many behavioral time-series (Huitema, 1985, 1986a). It is not highly unusual, however, to encounter autocorrelation with these designs. To rule out the presence of serial dependency, the residuals of the ANOVA model should be assessed. While software for conventional tests for autocorrelated errors of this model exist (they are known as the runs test and the Durbin-Watson test), recent research suggests that these tests are inappropriate in the case of samples of the size generally found in applied behavior analysis research. An alternative test proposed by Huitema and McKean (1994d) is both simple and effective for ANOVA and other variants of the general linear model applied to time-series data. A statistically significant outcome of this test implies that a more complex model should be considered.

Interrupted Time-Series Regression Models

If the residuals of the ANOVA model are significantly autocorrelated it is possible that within-phase trend in the data is responsible for the autocorrelation, in which case an appropriate interrupted time-series regression model should be identified. If this model is applied to the data and the residuals do not have significant autocorrelation, the conventional tests associated with the model are valid. It has been found that a general linear model which allows for a different slope in each phase of the experiment often fits behavioral data quite well (Huitema, 1985, 1988). Details on the regression approach for the analysis of time-series designs containing two, three, or four phases can be found in Huitema, McKean, and McKnight (1994). Once again, if the residuals of this approach are autocorrelated, some more complex alternative model should be considered. The form of the increased complexity may be in terms of added independent variables or parameters to model directly the errors of the model.

Generalized Least Squares Procedures

When the residuals of a time-series regression model are autocorrelated, a method of modeling the autocorrelation should be found. A well known approach for solving this problem is known as generalized least-squares. Two versions of this approach have been studied recently (McKnight, 1994); both were found to be clearly unacceptable in the case of small samples.

ARIMA Intervention Models

ARIMA intervention models (Box & Tiao, 1965, 1975; Glass et al., 1975) provide a general approach to describing time-series data. It is one of the approaches that is at the basis of the ongoing debate surrounding the statistical analysis of behavioral data. Many investigators recommend the use of ARIMA modeling over ANOVA or other conventional models for behavioral data because of the alleged autocorrelation problem (e.g., Hartmann et al., 1980). Some have attempted to explain ARIMA modeling in terms that can be readily understood by behavioral researchers to enable them to use ARIMA procedures effectively (e.g., Jones et al., 1977; Hartmann, et al., 1980). Unfortunately, ARIMA techniques are complex and much has been lost in the attempted translation to simplicity.

General ARIMA models are specified using three components (*p*,*d*,*q*); where *p* is the autoregressive (AR) component, *d* is a term that describes a maneuver called differencing that removes trend from the data set, and *q* describes the moving average (MA) component[1]. Three numbers specify the order of the three components. ARIMA (0,0,0) specifies the white noise model (Glass et al., 1975), where no autoregressive component is identified, no differencing is required to remove trends, and no moving-average component is inherent. Data that conform to the white noise model are not autocorrelated and any variation in the data is considered to be a random process. A first-order autoregressive model contains one AR parameter, but requires no differencing or moving average parameters. It is identified as an ARIMA (1,0,0). Complex models can contain autoregressive, differencing, and moving average components of high order (e.g., 3,2,2), but this is rare. The major tools for specifying the order (*p*,*d*,*q*) are the sample autocorrelation and sample partial autocorrelation functions (Box & Jenkins, 1976).

Problems in Applying ARIMA Intervention Models With Behavioral Data

There are at least three problems with attempting to apply ARIMA modeling to behavioral data: (1) there generally is an inadequate number of data points for accurate model identification, (2) differencing to remove trends from the data introduces both

---

[1] Moving averages is a complex phenomenon, better labeled as a "moving sum." The MA model describes the error at time *t* as a function of previous and current "random shocks" to the system. Most behavioral time series do not involve MA components, thus it is neither necessary nor desirable to attempt an explanation of moving averages here.

statistical problems and interpretation issues, and (3) a subjective component is involved in model identification. Each of these will be considered subsequently.

Small Samples

The foremost problem with applying ARIMA modeling to behavioral data is related to the number of data points typically available in applied behavioral studies. Box and Jenkins (1976) state that a minimum of 50 observations is necessary to obtain a stable estimate of the true autocorrelation function. The modal number of observations in baseline phases of studies published in *JABA* from 1968 to 1977 was about 4 (Huitema, 1985). Clearly this number comes nowhere near the recommended minimum. Large samples of time-series data points are needed "for knowledge of p, d, and q, so that the dependence among observations can be properly accounted for in statistical tests of intervention effects" (Glass et al., 1975, p. 112). Huitema (1988) clarifies:

> Since the identification of an ARIMA model depends upon the estimation of a substantial number (say a dozen) of reasonably stable autocorrelations, the $n = 50$ recommendation is sensible. It must be understood that a major purpose of the autocorrelation function in ARIMA modeling is the identification of an appropriate model. This is accomplished by comparing the estimated autocorrelation and partial autocorrelation functions with the autocorrelation and partial autocorrelation functions known to characterize various ARIMA models. (p. 271)

Jones et al. (1975) and subsequent investigators (Busk & Marascuilo, 1988; Sharpley & Alavosius, 1988; Suen, 1987; Suen & Ary, 1987) have failed to recognize the importance of model identification. Further, Sharpley (1987) has suggested that model identification may not even be necessary, stating that "the argument that accurate model-identification is difficult with short series is based on the assumptions that exact model-identification is both possible and necessary" (p. 41). He goes on to say that "it would appear that the need to be specific in model-identification is not so obvious as was once thought" (p. 41). This position is indeed peculiar when he later reports on a personal communication from Glass (November 10, 1984):

> The often-quoted figure of 50 to 100 observations as necessary for the reliable use of ITSA [interrupted time-series analysis] is a vast overestimate, and...as few as eight baseline and eight intervention observations may be sufficient to test for treatment effects, *providing the model of the series is known* [italics added] (p. 42).

Sharpley seems to be saying here that accurate model identification is difficult and unnecessary, but that in order to use interrupted time-series analysis accurately, the model of the series must be known.

Differencing to Eliminate Trend

Differencing is a procedure for removing trend from a data set. Glass et al. (1975) describe the procedure as follows:

> If the series is not stationary, successive differences are taken until the resulting series is stationary; for example, if the values of a series are 1, 3, 5, 7, 9, etc., there is a linear trend. First differencing reduces the series to 2 (=3-1) [sic], 2 = (5-3), 2 (=7-9) [sic], 2 (=9-7) [sic] and the resulting series of all 2's is clearly stationary. If the values of a series are 1, 4, 9, 16, 25, 36, there is a quadratic trend. First differencing gives 3, 5, 7, 9, 11, and differencing again gives 2, 2, 2, 2. In general, first differencing eliminates linear trend, second differencing eliminates quadratic trend, and so on (p. 79).

While trend in a baseline data set may mask intervention effects under some circumstances (e.g., where the trend is in a therapeutically desirable direction), behavior analysts do not typically recommend statistically removing the trend. Rather, the preferred option is to wait for the data to stabilize before further experimental manipulations are initiated. Sidman (1960) asserts that when an experiment produces unstable behavior, the investigator must rule out extraneous variables that may influence the dependent measure. Once the experimenter is assured that the instability is a product of the experimental manipulation, the next task is to "examine the instability and describe it as fully as his [or her] available methods will allow" (p. 271). The rationale for this recommendation is that accurate description of the form of the instability and its relevant controlling variables is necessary if subsequent experimenters are to successfully deal with the phenomenon as it arises.

Kazdin (1982) states that "a problem may emerge, at least from the standpoint of the design..." (p.263) when trend in a therapeutically desirable direction is apparent in baseline data. In this case it may be difficult to detect intervention effects. He presents several design alternatives for dealing with trending data, including variants on the reversal, multiple-baseline, and changing-criterion designs. The differential effects of treatment can often be detected under these design options, even when there is overall trend in the data. In cases where trending overshadows or makes ambiguous the intervention effects, a statistical aid may be employed.

Subjective Nature of ARIMA Model Identification

Even with large sample sizes, ARIMA model identification is complex, time consuming, and inefficient, especially for "mathematically unsophisticated researchers" (Velicer & McDonald, 1984, p. 39). Human judgement is required at several stages in the procedure, and there is "no reason to assume that given a standard set of data, all persons would arrive at exactly the same model" (Horne, Yang, & Ware, 1982, p. 187). Michael's (1974) advice regarding the use of complex inferential techniques is skeptical:

> The various judgemental aids do not achieve their simplifying effects without some cost...they are easier to react to in part because they are abbreviations...Further, the scientist must spend some time learning about them, time he [or she] might be spending in other activities relevant to his [or her] subject matter (p. 647).

The amount of effort required to gain the necessary repertoire for accurate model identification, were it possible, is significant. The cost of such complex procedures, to both the behavioral researcher and the research consumer, in terms of time and effort might greatly outweigh any benefit accruing from their use. Recent developments in automated model identification algorithms (Liu, 1993) may reduce this objection to ARIMA modeling . It is obvious from the current work on such algorithms, however, that they are not designed to be used with sample sizes that are typical in behavioral psychology. The shortest series in a large collection of recent examples demonstrating automated identification (Liu, 1993) had 70 observations.

New Double Bootstrap Procedure

The poor small-sample performance of the conventional versions of the generalized least squares procedure and several problems that accompany ARIMA modeling motivated the development of a new approach. This approach employs a computer intensive algorithm known in statistics as the "bootstrap" during two stages of the analysis. The original evaluation of the application of this method to the analysis of the interrupted time-series design reported very satisfactory results (McKnight, 1994). Current work confirms this evaluation (McKnight, McKean, & Huitema, in preparation).

Benefits of GLM and New Double Bootstrap Models Over ARIMA Modeling

There are at least two benefits associated with using the conventional general linear model and the new double bootstrap methods when analyzing behavioral data. First, they are not bound by large-sample requirements in the same manner as are the ARIMA techniques. Second, they produce results that are easier to understand than those produced through ARIMA modeling. Each of these points will be considered below.

Solution to Small Sample Problems

It has been repeatedly found that a relatively large N (at least 50) is required for adequate estimation of autocorrelation parameters and model identification when employing either ARIMA procedures (e.g., Box & Jenkins, 1976; Horne et al., 1982; Huitema, 1985, 1986a, 1988; McKnight, McKean, & Huitema, in preparation) or generalized least squares (McKnight, 1994). Further, it is evident that the number of observations typically generated in studies published in *JABA* comes nowhere near the recommended minimum of 50. Rather, they typically contain fewer than ten observations per phase (Huitema, 1985).

While small sample sizes are often an issue when employing conventional general linear models such as ANOVA, the concern is one of power rather than of the stability of model identification and parameter estimation; that is, there is a potentially low probability of concluding there is an experimental effect when in fact one has occurred. However, the power of an inferential test is also related inversely to the amount of within-phase variation. Thus, where experimental control is effective and data show little variability within phases, the power of tests based on the general linear model can be relatively high even in small-N cases.

In the case that the errors of the general linear model are autocorrelated, the new bootstrap procedure mentioned above has been shown to be effective with small-N data sets. This procedure enables the researcher to accurately estimate autocorrelation parameters and intervention effects without relying on large-sample theory that is of questionable relevance in the case of small samples.

Type III Errors

Behavior analysts generally favor the most direct approach to data analysis: visual inspection of the graphed data to determine if intervention effects are evident. This approach leaves little room for speculating on which aspects of the stimulus array generated by graphed behavioral time-series data control the investigator's conclusion that an effect did or did not occur. Said another way, when we visually inspect data, we can point to the parts of the graph that have led us to our conclusions regarding the effectiveness of the treatment. Similarly, when conventional general linear models such as ANOVA are used, the single-subject researcher can easily point to the relevant aspects of the stimulus array to explain what the significance test is comparing. Unfortunately, behavioral researchers are now encouraged to use software that will fit complex ARIMA intervention models. A problem associated with processing one's data through such software is the high likelihood of misunderstanding the nature of the comparison generated. Huitema (1986b) calls this a Type III error. He points out that the output generated by complex time-series intervention analysis (ARIMA) programs contain terms and probability values that are likely to answer questions that are very different from the question the researcher thinks is being answered.

Summary of the Major Methodological Problems Associated With Previous Surveys of Autocorrelation in the *Journal of Applied Behavior Analysis*

Many misunderstandings regarding autocorrelation and time-series analysis have been published in the behavioral literature. A detailed account of most of these issues can be found in several critical articles (Huitema, 1985, 1988) and chapters (Huitema, 1986a, b). The argument underlying all published recommendations to use ARIMA (and other) complex time-series methods in the analysis of applied behavior analytic data rests on the presumption that the errors of such studies are autocorrelated.

This implies knowledge of the distribution of appropriate autocorrelation coefficients associated with a population of applied behavior analytic data. All previous attempts to characterize this distribution have been flawed. The most important flaws associated with each of the earlier attempts have been presented by Huitema (1990). He labeled these errors as (a) model misspecification type I (the problem of the 1970s), (b) model misspecification type II (the problem of the 1980s), and (c) the universal estimation problem; these are summarized in the remainder of this section.

The first problem is associated with the work published in the 1970s; an early study by Jones et al. (1977) involved the computation of autocorrelation coefficients on 24 sets of raw data from studies published in the *Journal of Applied Behavior Analysis* (*JABA*). The authors concluded that 83% of the studies had significant high levels of autocorrelation and that ARIMA time-series models were required for the analysis of these studies. The principal recommendation of the Jones et al. (1977) paper that complex time-series analyses be used as a supplement to visual analysis of behavioral data is questionable because it is based in irrelevant autocorrelations. The error committed in this work can be viewed as a type of model misspecification. The model the researchers used in conceptualizing their analyses was not formally stated in their work, but verbal descriptions of their methods, the obtained autocorrelation coefficients, and subsequent reanalyses of their data leave no ambiguity whatsoever that the following statistical model captures the strategy they used:

$$\text{(1970s Model)} \qquad Y_t = \mu + \varepsilon_t. \qquad (1)$$

This model implies that behavior is a function of μ, the overall level of the whole observed series (i.e., data collected across all conditions or phases), and ε_t, random error. It is known that this model misspecifies the actual form of the behavior in any time-series in which there is a difference in level between phases (Huitema, 1985, 1986a). Because the model does not contain a term that acknowledges differences in the levels of the different phases, the model is said to misspecify the sources of variation in the experiment. This means that if there are different levels of behavior in the different phases, artificial autocorrelation will necessarily be introduced to the errors.

While the authors of the work carried out in the 1970s claimed that their autocorrelation results were proof that data analysis methods based on the general linear model (such as independent sample t-tests and ANOVA F-tests) were inappropriate for behavioral time-series data, their results were shown to be completely irrelevant to such models. The autocorrelation coefficients computed under the 1970s model provide no information on the autocorrelation associated with the errors of any realistic intervention model (Huitema, 1986a). For example, if the ANOVA model describes the data, the residuals for a study with two or more phases are based on deviations from each phase mean rather than from one common mean for the entire series. If the raw data are inappropriately combined into a single series, a high positive autocorrelation coefficient will generally be obtained, even if the residuals of the ANOVA model have zero or negative autocorrelation (Huitema, 1985, 1986a). Unfortunately several subsequent studies appeared, also claiming that formal time-series analysis of behavioral data is not only desirable but necessary (e.g., Sharpley & Alavosius, 1988; Suen, 1987; Suen & Ary, 1987).

The problem just described was acknowledged in the subsequent surveys of autocorrelation in behavioral data carried out in the late 1980s and early 1990s by Busk and Marascuilo (1988) and Matyas and Greenwood (1985, 1991). These researchers also failed to explicitly describe their model, but descriptions of their work imply the following:

$$\text{(1980-1990s Model)} \quad Y_{tj} = \mu_j + \varepsilon_{tj}. \tag{2}$$

This model is much more realistic because it is a representation of the behavior that occurs within a phase rather than across all phases. Autocorrelations computed on the errors of this model are meaningful if the data are not trending upward or downward across time. If the data from two or more phases are conceptualized with one model rather than as a collection of individual phase models, it can be written as

$$Y_{tj} = \mu + \alpha_j + \varepsilon_{tj}. \tag{3}$$

This is the ANOVA model for a time-series design containing multiple phases. The term μ is the overall level of the whole series (across all phases), α_j is the effect of being exposed to condition j in phase j, and ε_{tj} is the error. If the errors of model 2 for each individual phase in a multiple phase design are independent, the errors for the ANOVA model (3) associated with this design will also be independent. The problem with models (2) and (3) is that they are unrealistic in the case of data that trend within phases. This is another case of misspecification error. Trends within phases will introduce artificial autocorrelation to the errors of models (2) and (3) just as differences between levels of phases introduces artificial autocorrelation in model (1). Both level changes between phases and trending within phases are deterministic aspects of the data; such characteristics should be built into the model of behavior. If the model does not contain parameters to explain these characteristics when they exist, the model is misspecified and the errors will be artificially autocorrelated. The models recently studied by Huitema, McKean, and Zhao (in press) contain parameters for both differences in level and trend for these very reasons.

The third error described by Huitema is called the universal estimation problem. All of the previously published surveys of autocorrelation in behavioral data employed the conventional autocorrelation estimator. This estimator is known to yield biased estimates. The degree of bias has been shown to be serious in the case of small samples (Huitema & McKean, 1991). This implies that all previous surveys are flawed in the sense that the measure of autocorrelation universally used in these studies is not adequate for the purpose of characterizing the true level of autocorrelation in the field of applied behavioral research. The major purpose of the present research is to provide an unbiased estimate of the level of autocorrelation present in the errors of the population of recently published behavioral studies.

METHODS

Introduction to Meta-analysis

In 1985 Huitema concluded that the residuals of general linear models fit to applied behavioral data have little or no autocorrelation. Given that the typical baseline phase in studies published in *JABA* is around four to six observations, one might question the means by which Huitema came to this conclusion. The approach was to cumulate the results from several hundred applied behavioral data sets in order to summarize the overall level of autocorrelation in the population of behavioral studies. The general approach taken by Huitema (1985) and in the current study is commonly referred to as "meta-analysis."

Logic of Meta-analysis

Meta-analyses are statistical methods for combining the results of a collection of individual studies. The logic behind these procedures is as follows:

> [If each] computed effect size from a series of studies estimates a single population effect size, then it is sensible to compute a single pooled estimate that will be far more precise than any of the individual estimates, and the corresponding tests of significance will be substantially more powerful than any of the separate tests (Raudenbush, 1991, p. 242).

Two important features of meta-analytic procedures relevant to the current study, and those of Huitema (1985, 1986a, 1988), are that they enable the researcher to (1) summarize an overall relationship between two variables by cumulating the results of independent studies, and (2) determine which factors, called moderating variables, are associated with variations in results across individual studies (Rosenthal, 1991). A brief overview of the steps taken in conducting a meta-analysis follows.

General Steps in Meta-analyses

The starting point in a meta-analysis is to bring together a collection of studies from a common area, each with their individually computed results. Johnson, Mullen, and Salas (1995) observed that the results of empirical research are typically reported in terms of effect size (i.e., the magnitude or strength of the effect) or significance level (i.e., the likelihood of having obtained the observed results given that the null hypothesis of no effect is true). Several metrics are available for estimating and reporting effect size; three commonly employed methods are Pearson's *r*, the transformation of Pearson's *r* to Fisher's Z, and the standardized mean difference, sometimes known as Cohen's *d*. Similarly, various methods are available for reporting significance levels; commonly used metrics include standard normal deviates (*z* scores) and probability (p) values (Johnson et al., 1995). It is conceivable that a collection of studies from a common area may employ multiple means of measuring the behavior of interest and reporting on effect sizes and significance levels. It is not necessary for the dependent measures of these studies nor their corresponding summary statistics to be in the same metric in order for the results to be meaningfully combined (see Rosenthal, 1991, chapter 2).

Because meta-analyses include statistical tests of many individual analyses, evidence for the appropriateness of pooling the individual results is needed. The consistency of study results can be assessed with a chi-square test of homogeneity of the individual effect size coefficients (Huitema, 1988; Raudenbush, 1991). A test of the hypothesis of homogeneity with a nonsignificant outcome is evidence that the results of the independent studies are indeed based on samples taken from a common population. Thus, pooling the results is justified. The pooled coefficient is an optimal estimate of a common population effect that is weighted by sample size. Those individual coefficients associated with larger sample sizes are given more weight than are those associated with smaller samples.

Once the pooled weighted estimate is computed, it is then reasonable to construct confidence intervals around the pooled estimate and perform significance tests against the null hypothesis of no effect in the population. Tests of significance on the pooled estimate will be considerably more powerful than any of the tests on the results of the individual studies.

Identifying Moderator Variables

If the hypothesis of homogeneity is rejected, the next step in the meta-analysis is the search for moderator variables: factors independent of the treatment that might affect the outcome of a study. Rosenthal (1991, p. 9) provides an example of how moderator variables can be useful in explaining variation in effect size across several studies. In his example, the effect size of interest is the correlation between teacher expectations and pupil performance in six separate studies. Assume that a test of the homogeneity of the six correlation coefficients was conducted and the p-value produced was small (i.e., $p \leq .05$); the hypothesis of homogeneity would be rejected and we would have no evidence to support pooling those coefficients. Rosenthal provides a second set of numbers, mean ratings of teacher excellence, that might provide an explanation for the inconsistent correlation coefficients found in the six studies. Assume also that a significant relationship between the original effect size estimates (i.e. the correlation between teacher expectations and pupil performance), and the moderator variable (mean ratings of teacher excellence) exists. We could empirically demonstrate that the level of teacher excellence moderates the magnitude of the relationship between teacher expectations and pupil performance, concluding that the results of the six studies are consistent when the moderator variable is considered in the statistical model assumed to explain the data.

Current Development of Meta-analysis Methods for Autocorrelation Coefficients

A meta-analysis of the degree of autocorrelation in typical applied behavior analytic data is possible using the procedures outlined above with slight modifications. For instance, instead of a collection of effect size coefficients, we have a number of autocorrelation coefficients computed on observations from several behavioral time-series data sets. Because autocorrelation coefficients are not generally found in studies published in applied behavioral journals, the appropriate steps in such a meta-analysis would begin with: (a) sampling a number of time-series data plots from published applied behavioral studies, (b) extracting the quantitative data from those plots, (c) computing individual autocorrelation coefficients for each sampled data set, and (d) computing test statistics on the amount of variation away from zero observed in the distribution of the autocorrelation estimates. A test of the homogeneity of the autocorrelation coefficients--a modified version of a chi-square test designed specifically to assess autocorrelation coefficients--would follow. Depending on the results of the test of homogeneity, one would then either (a) compute a pooled weighted mean autocorrelation estimate (given nonsignificant results from the homogeneity test) and construct confidence intervals around that estimate or (b) search for moderator variables that might account for the apparently inconsistent results of the individual studies and include information on those moderator variables in the appropriate statistical model fit to the data. This was the general routine followed in the current study. Details of sampling procedures and meta-analytic procedures modified for use with autocorrelation coefficients can be found in the sections to follow.

Data Acquisition Procedures

Two primary samples of data were used in this study; one newly sampled from *JABA* and one obtained from Busk and Marascuilo (1988). To remain consistent with previous attempts at identifying autocorrelation in behavioral data (e.g., Busk & Marascuilo, 1988; Huitema, 1985; Sharpley & Alavosius, 1988; Suen & Ary, 1987), the analysis was restricted to data sets with a minimum of six observations within the first phase of the design. By limiting the data to those meeting this phase length criterion, there were no concerns regarding the properties of the autocorrelation estimators or the associated standard errors. Also, issues regarding the distortion of autocorrelation estimates caused by deterministic intervention effects were avoided by focusing on data obtained before interventions were applied.

Sampling Methods for Selecting Data Sets From *JABA*

The period covered in *JABA* from volume 23, issue 1 (1990), through volume 28, issue 1 (1995) was the source of 200 time-series data sets employed in this study. First baselines, defined as the first phase of data collected under baseline or functional analysis conditions prior to any intervention or subsequent conditions, were sampled. Thus, the first A of an ABA design would qualify for inclusion, but the second A (i.e., the reversal or withdrawal phase) would not. Similarly, the first phase of a pairwise test-control design involving multiple assessment phases (Iwata, Dorsey, Slifer, Bauman, & Richman, 1994) would qualify for inclusion but the second and subsequent phases would not. Data sets plotted with some measure of time on the abscissa were included in this finite population. Excluded were cumulative graphs, bar graphs, histograms, and other graphs not plotted as a time series.

Two subcategories of "first baseline" were defined prior to sampling; (1) functional analysis baselines, and (2) baselines from conventional reversal, multiple-baseline and other designs. Functional analyses consisted of conditions in which target behaviors were observed across two or more well defined environments. Data obtained under different functional analysis conditions are plotted on separate data paths and two or more data paths may be presented for a single assessment phase. Under the functional analysis paradigm, physical and social aspects of the environment that might affect the occurrence of a target behavior are manipulated and a base rate of behavior under these conditions is observed (Iwata et al., 1994). Functional analysis baselines are a homogenous group of conditions in which the behavior of interest is typically some form of self-injurious behavior (SIB) or stereotypy, although the latter is studied less often using this assay. Conditions typically employed to assess behavior include social-positive reinforcement, social-negative reinforcement, automatic reinforcement and control conditions (see Iwata, Pace, et al., 1994 for examples). Of the 495 functional analysis first baselines appearing in the *JABA* issues sampled, 260 contained six or more observations; 50 of these were randomly selected for inclusion in the meta-analysis.

Baselines from reversal, multiple-baseline and other designs consist of a heterogeneous collection of first phases in which some behavior is measured over time under existing environmental conditions. These conditions are free to vary throughout the baseline prior to any intervention. Thus, these baselines contain data from a wide variety of subjects and behaviors and may be based on individual subjects, larger molar units including well defined (stable) groups, or dynamic groups such as the general public in community-based interventions. Of the 1,955 data sets found in this category, 1,173 contained six or more observations; 150 of these were randomly selected for inclusion in the meta-analysis.

Data sets that were selected as the original sample of 50 functional analysis series or as the original sample of 150 conventional baselines were excluded from these samples if all observations fell at the minimum or maximum of the relevant metric (e.g., all at 0% or all at 100%). If such a data set was selected, it was discarded and a new set was drawn randomly[2].

Data Extraction Procedures

Quantitative data were extracted from the behavioral time-series graphic data displays using the computer program DataThief (Huyser & van der Laan, 1992). DataThief, developed to aid nuclear physicists in the reverse engineering of data plots, allows the user to generate raw data from existing graphic displays. Graphic displays of sampled data were scanned using a Hewlett Packard ScanJet IIcx and WordScan™ software. Each graphic image was then saved in a PICT format and subsequently opened into the DataThief program. When opened into DataThief, the data plot appears on the screen and the program prompts the user to; (a) point the cursor and click the mouse once on each end of the horizontal axis, (b) enter the values corresponding to each end of the axis, (c) point the cursor and click the mouse once on each end of the vertical axis, (d) enter the values corresponding to each end of the axis, and (e) move the cursor to the first data point to be read and click the mouse once, and point and click on each subsequent data point. When the user has clicked on all the data points, the values are saved as an ASCII file that can be read by any spreadsheet or graphics program. Generated data were imported to Minitab statistical software for analysis.

---

[2] From here on, the term "data sets" will refer to the individual time series contained in each sample, and the term "sample" will refer to the groupings of time-series data sets subsumed under the 1990-1995 *JABA* sample, the conventional baseline and functional analysis subsamples and the Busk and Marascuilo sample.

Reliability of Data Extraction Procedures

Twenty data sets were randomly selected from the 1990-1995 *JABA* sample (N=200) to be used for estimating the reliability of data extraction procedures. The PICT formats for these 20 data sets were opened into the DataThief program and the data were extracted a second time following the procedures outlined in the previous section. The conventional autocorrelation estimator r_1 was computed for the data extracted from the 20 reliability data sets and for their corresponding 20 original sample data sets using the formula presented on page 13 (Computational Basis of Autocorrelation), where k=1. A Pearson correlation coefficient was computed between the r_1 values associated with the 20 reliability and 20 original extracted data sets; the correlation between these two sets of autocorrelation coefficients was .998.

Busk and Marascuilo (1988) Data

The second sample of data employed, that used by Busk and Marascuilo (1988), was obtained from the authors for reanalysis. Busk and Marascuilo sampled 248 time-series data plots containing six or more observations that were published in *JABA* from 1975 to 1985. Of these 248 data sets, 101 first baselines (i.e., the first baseline in an experiment) were included in the current reanalysis (Busk & Marascuilo, 1988).

Computation of Summary Statistics

The computations described below refer to "raw data" and "linear detrended data." Because only single phase baseline data sets were used, applying the formulas to the "raw data" are equivalent to applying them to the residuals of the following linear model: $Y_t = \mu + \varepsilon_t$. "Linear detrended data" are the residuals obtained by regressing the raw data on time in order to remove trend. That is, they are the residuals that estimate the errors in the following model: $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$. All subsequently described descriptive and test statistics were computed on four samples of data; (1) the 200 newly sampled *JABA* (1990-1995) first baseline data sets, the subsamples of (2) 150 baselines and (3) 50 functional analysis baselines (separated out of the combined sample of 200), and (4) the 101 first baselines obtained from the Busk and Marascuilo (1988) sample.

Conventional Autocorrelation Estimator r_1 Applied to the Raw Data

Conventional lag-1 autocorrelation estimates r_1 were computed on the raw data (Y) of each series. This estimator is known to be negatively biased (Huitema & McKean, 1991, 1994b). If, for example, the process autocorrelation is zero, the expected mean of a frequency distribution based on a large number of sample autocorrelation coefficients would be expected to equal $-\frac{1}{n}$. It is important to recognize the negative bias inherent in this estimator; the autocorrelation coefficients from the individual studies provide the foundation for subsequent meta-analyses and decisions based on those analyses. The r_1 formula is written

$$r_1 = \frac{\sum_{t=1}^{n-1}(Y_t - \overline{Y})(Y_{t+1} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$$

This is the formula most frequently presented in time-series analysis textbooks and most often used in widely available statistical software packages for computing autocorrelation coefficients on time-series data.

Modified Autocorrelation Estimator r_{F1} Applied to the Raw Data

A newly developed autocorrelation estimator r_{F1} (Huitema & McKean, 1994b) was the second coefficient computed on the raw data of each series. The coefficient r_{F1} was developed to avoid the problem of bias found in the conventional estimator. The r_{F1} formula is written

$$r_{F1} = \left[r_1 + \frac{1}{n}\right]\left[1 + \frac{5}{n}\right]$$

where r_1 is the conventional lag-1 autocorrelation coefficient applied to the original series (prior to detrending) and n is the number of observations upon which r_1 is based.

Huitema and McKean (1991) demonstrated that the amount of bias in the r_1 estimator is a function of the true value of the autocorrelation in the process (ρ), and the sample size. The estimator r_1 is highly biased in the case that the process autocorrelation ρ is a high positive value, especially when based on sample sizes that are typical of applied behavioral studies, whereas r_{F1} is almost unbiased with most levels of process autocorrelation, even with small n (Huitema and McKean, 1994).

Conventional Autocorrelation Estimates Applied to Linear Detrended Data

A second source of bias that artificially inflates autocorrelation estimates is present in many applied behavioral studies; trend in the data series introduces a positive bias into the already biased r_1. It is expected that when the data are appropriately detrended the autocorrelation distribution will be similar to that expected by chance when sampling from a population of independent (nonautocorrelated) errors. To detrend the data, the raw data were regressed on time. The conventional autocorrelation estimate computed on the detrended data is denoted r_{1R}.

Reduced Biased Autocorrelation Estimator Applied to Linear Detrended Data

The reduced bias r_{F1} estimator was developed for use with raw scores. A corresponding unbiased estimator for use on the residuals of linear detrended data is not currently available, and little is known about the properties of the r_{F1} estimator when applied to the residuals of a simple linear regression. Further, a corresponding test of significance of r_{F1} applied to detrended data is not available. Thus, an estimator developed by Huitema and McKean (1995) for use in testing for autocorrelated errors in time-series regression models was employed. This estimator is defined as follows:

$$r_{HM} = \left[r_{1R} + \frac{P}{n} \right]$$

where r_{1R} is the r_1 estimator computed on the detrended data, P is the number of parameters in the model, and n is the number of observations upon which the coefficient is based. The value of P in the case of the simple linear regression model used for detrending is two. This estimator is essentially unbiased when sampling from error processes having autocorrelation values near zero.

t-values on Slope as a Standardized Descriptor of Trend in Sample Data Sets

When the raw data are regressed on time for each individual data set in order to remove the influence of trend from the autocorrelation analysis, a *t* value can be obtained to test the null hypothesis that the population slope is zero; it may also serve a second, descriptive, function. The *t* statistic is computed by dividing the sample slope coefficient by an estimate of its standard error as follows:

$$t = \frac{b_1}{\sqrt{\frac{MS_{Residual}}{\Sigma x^2}}}$$

where b_1 is the sample slope, $MS_{Residual}$ is the mean square residual obtained from the analysis of variance of the regression of Y on time, and Σx^2 is the corrected sum of squares on the time variable. The *t* values resulting from performing the regression procedure on each data set contained in the samples were plotted as frequency distributions. These distributions provide descriptions of the standardized slope in the data sets contained in the samples. The problem with the raw slope as a measure of trend is that they are not all based on the same metric (on either the time dimension or the dependent variable). Dividing the slope by the estimated standard error of the slope standardizes the b_1 values and hence converts them to a common metric.

Inferential Procedures

Test Statistic for the r_{F1} and r_{HM} Autocorrelation Coefficients

The test statistic associated with the reduced bias autocorrelation estimator (r_{F1}) for the raw data series was computed using the formula presented in Huitema and McKean (1994b); this statistic is interpreted as a standardized normal deviate and is defined as:

$$z_{F1} = \frac{r_{F1}}{\sqrt{\left(1 + \frac{5}{n}\right)^2 \left[\frac{(n-2)^2}{n^2(n-1)}\right]}}$$

where

$$r_{F1} = \left[r_1 + \frac{1}{n}\right]\left[1 + \frac{5}{n}\right]$$

Similarly, the test statistic associated with the estimator r_{HM} for the detrended data was computed using the formula which yields a standardized normal deviate:

$$z_{HM} = \frac{r_{HM}}{\sqrt{\left[\frac{(n-2)^2}{n^2(n-1)}\right]}}$$

where

$$r_{HM} = \left[r_{1R} + \frac{2}{n} \right]$$

Chi-square Tests of the Homogeneity of r_{F1} and r_{HM} Autocorrelation Coefficients

A chi-square statistic developed by Huitema, McKean, and Zhao (in preparation, c) was used to test the homogeneity of the autocorrelations. This statistic tests H_0: $\rho_1 = \rho_2 = \ldots = \rho_J$; that is, it is a test of the hypothesis that the process autocorrelation is the same in all J studies. If the null is retained, this is evidence that the variation of the sample autocorrelation coefficients around the central value may be due to sampling error alone (i.e., the sample *r* coefficients were drawn from a population of studies having the same degree of process autocorrelation). If the coefficients are considered homogeneous it is appropriate to pool them. If the null is rejected, then factors influencing the variation away from the central value (i.e., moderator variables) need to be found and included in the statistical model describing the data.

The chi-square test for the homogeneity of the r_{F1} coefficients was computed as follows:

$${}_{F1}\chi^2 = \sum_{j=1}^{J} \left[\frac{\left(r_{F1j} - \bar{r}_{F1w}\right)^2}{\left(1 + \frac{5}{n_j}\right)^2 \left[\frac{(nj-2)^2}{n_j^2\left(n_j - 1\right)}\right]} \right]$$

where ${}_{FI}\chi^2$ is the test statistic, r_{F1j} is the autocorrelation estimate based on sample j, J is the number of sample data sets contributing to the chi-square statistic, and n_j is the number of observations in the *j*th data series. This test statistic is evaluated against the critical value of the chi-square distribution based on J-1 degrees of freedom.

The mean of the r_{F1} coefficients weighted by sample size, $\bar{r}_{F1_w}$, was computed as

$$\frac{\sum_{j=1}^{J}\left[n_j \left(r_{F1j} \right) \right]}{\sum_{j=1}^{J} n_j}$$

Similarly, the chi-square test of the distribution of r_{HM} coefficients was computed as follows:

$$_{HM}\chi^2 = \sum_{j=1}^{J}\left[\frac{\left(r_{HMj} - \bar{r}_{HMw}\right)^2}{\left[\frac{\left(n_j - 2\right)^2}{n_j^2\left(n_j - 1\right)}\right]}\right]$$

where ${}_{HM}\chi^2$ is the test statistic, r_{HMj} is the autocorrelation estimate associated with the *j*th data set, J is the number of data sets in the sample contributing to the chi-square statistic, and n_j is the number of observations in the *j*th data series. This test statistic is evaluated against the critical value of the chi-square distribution based on J-1 degrees of freedom.

The mean of the r_{HM} coefficients weighted by sample size, $\bar{r}_{HMw}$, was computed as

$$\frac{\sum_{j=1}^{J}\left[nj \left(r_{1Rj} + \frac{2}{n_j} \right) \right]}{\sum_{j=1}^{J} n_j}$$

95% Confidence Interval on ρ Based on the Pooled Weighted Mean Autocorrelation Coefficient $\bar{r}_{HM}$

Confidence intervals around ρ_1 based on the pooled weighted mean $\bar{r}_{HM_w}$ were computed as follows:

95% Confidence Interval on ρ_1:

$$\bar{r}_{HM_w} \pm \left[\sqrt{\frac{1}{\Sigma n_j}}\,(1.96)\right]$$

where $\bar{r}_{HM_w}$ is the pooled weighted autocorrelation for the *j*th group and Σn_j is the sum of the number of observations contained in the individual data sets contributing to the weighted mean.

Population Variance Estimates of ρ_j and ε_j

It is recognized that sampling error, expected in any coefficient based on sample data, is exacerbated in small-n cases. The model used in this dissertation to explain autocorrelation in sample data using, for example, the r_{F1} estimator is

$$r_{F1j} = \rho_j + \varepsilon_j$$

where r_{F1j} is the sample lag-1 autocorrelation coefficient based on the *j*th process, ρ_j is the lag-1 autocorrelation associated with the *j*th process, and ε_j is sampling error $\left(r_{F1j} - \rho_j\right)$. It is assumed that the errors are normally and independently distributed around zero. Thus, the sample autocorrelation can be conceptualized as the sum of two components: the true but unknown process autocorrelation plus sampling error (Huitema, 1988). This statistical model explains why variation in sample autocorrelation coefficients can be anticipated when the autocorrelation (ρ_j) is constant across different processes.

The model does not explain the total expected variation in autocorrelation coefficients when they are sampled from processes with differing ρ_j values. In such a case, more terms are needed in the statistical model because sample autocorrelation coefficients may vary due to both sampling error and differences among the process autocorrelations. This model is written:

$$r_{F1j} = \bar{\rho} + \left(\rho_j - \bar{\rho}\right) + \varepsilon_j$$

where r_{F1j} is the sample lag-1 autocorrelation coefficient based on process j, $\bar{\rho}$ is the mean of the process autocorrelations, and ε_j is sampling error $\left(r_{F1j} - \rho_j\right)$ associated with r_{F1j}. Thus r_{F1j} is the sum of three components: the overall mean of the process autocorrelation coefficients, the difference between ρ_j and $\bar{\rho}$, and sampling error (Huitema, 1988).

Huitema (1988) proposed three essential questions based on the above three-component model: (1) What is the level of the overall process autocorrelation? (2) What is the variance of the lag-1 process autocorrelations? and (3) What is the variance of the sampling errors $\left(r_{F1j} - \rho_j\right)$? He contended that the answers to these questions provid a basis for the appropriate interpretation of obtained autocorrelation distributions. The *z* tests described in the previous section address the question of the overall level of the process autocorrelations by testing the obtained r_{F1} values against the postulated $\rho = 0$. Of particular relevance to this section is the difference between the two theoretical variances that explain the variance in the observed distribution of autocorrelation coefficients. The partitioning of the variance can be modeled as

$$\sigma^2_{r_{F1}} = \sigma^2_{\rho_j} + \sigma^2_{\varepsilon_j}$$

where $\sigma^2_{r_{F1}}$ is the variance of the population distribution of r_{F1} values, $\sigma^2_{\rho_j}$ is the variance of the process autocorrelations ρ_j, and $\sigma^2_{\varepsilon_j}$ is the variance of the errors. If all

r_{F1} coefficients are sampled from a single process, then $\sigma^2_{r_{F1}}$ will be equal to σ^2_{ε}. Thus if we compute estimates of $\sigma^2_{r_{F1}}$ and σ^2_{ε}, the difference between these two estimates will give a reasonable estimate of the remaining variance due to differences among process autocorrelations.

RESULTS

Results regarding the characteristics of the four samples of behavioral time-series data sets are described in this section. A description of the length of the many time-series data sets is presented first. Next, results on the distributions of various measures of autocorrelation are described in detail. Last, the characteristics of the observed distributions are compared with relevant theoretical distributions.

Descriptive Statistics

Length of Time-Series Data Sets

The length of the time-series data sets contained in the four samples is illustrated in Figure 4. Each distribution in the figure illustrates the number of observations included in each time-series data set. Summary statistics based on the distributions are presented in Table 3. The 1990-1995 *JABA* sample (N=200) has a slightly lower mean number of observations per data set than does the sample of 101 first baselines employed by Busk and Marascuilo (1988), but the subsample mean of 150 conventional baselines is almost identical to the Busk and Marascuilo sample. The functional analysis subsample (N=50), however, is characterized by a much lower mean number of observations per data set and a narrower range. The maximum series length of the subsample of conventional baseline data sets (N=150) and the 101 Busk and Marascuilo baselines are 106 and 69 observations, respectively; the maximum length of the functional analysis subsample (N=50) is only 17 observations. It should be kept in mind that one criterion for selection of these time-series data sets was that they contain a minimum of six observations. There were 495 functional

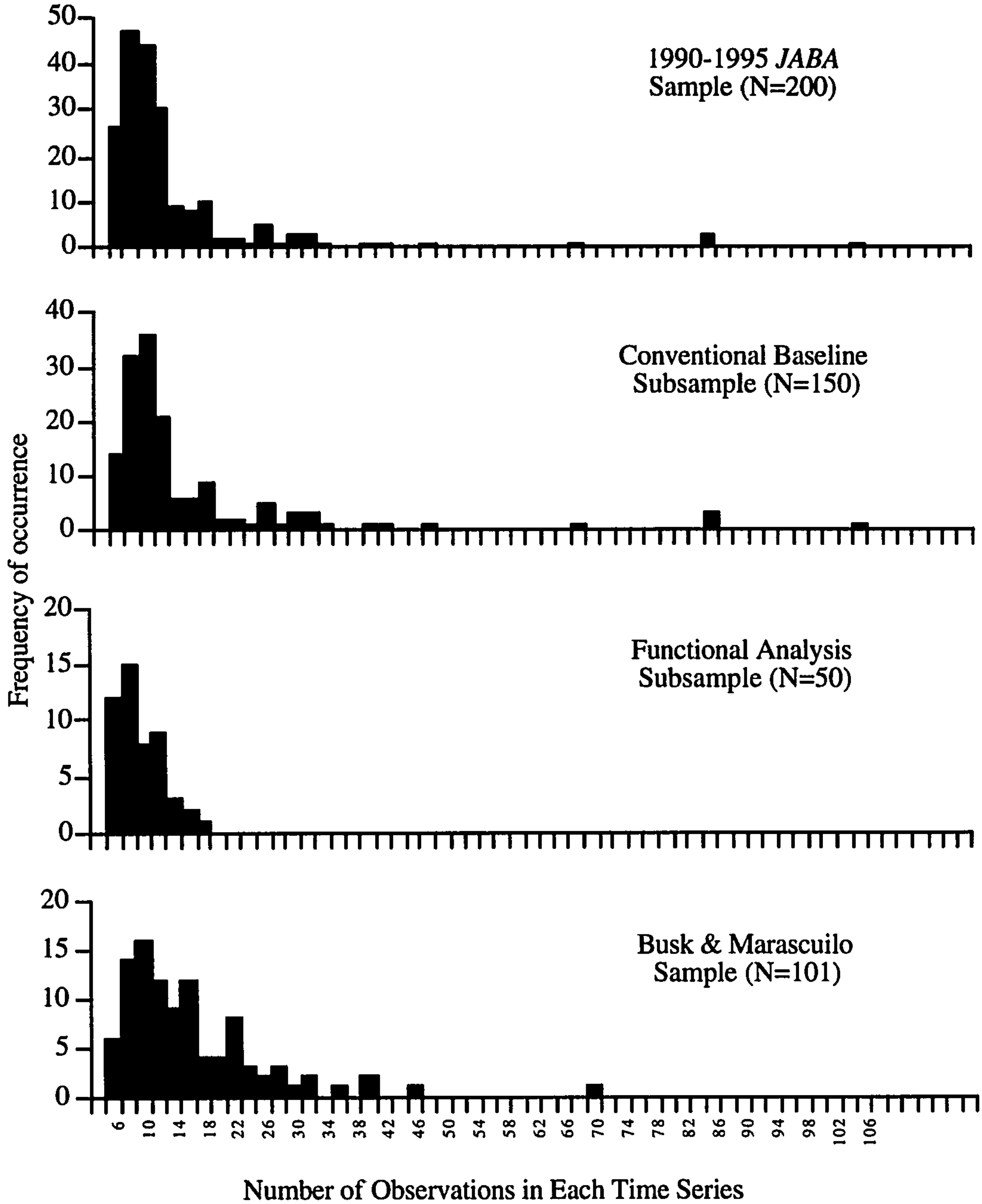

Figure 4. Frequency Distributions for the Number of Observations in Four Samples of Applied Behavioral Data Sets (Censored for n < 6 Observations).

Table 3

Number of Data Sets, Mean and Median Number of Observations, and Range of Values in Each Sample Contained in the Study

Sample	Number of data sets contained in the sample (*N*)	Mean number of observations (*n*) per data set	Median number of observations per data set	Range of values in each sample (*N*)
1990-1995 *JABA* sample	200	13.69	8	6-106
Conventional Baseline Subsample	150	15.21	10	6-106
Functional Analysis Subsample	50	9.12	10	6-17
Busk & Marascuilo Sample	101	15.75	13	6-69

analysis and 1,955 conventional baseline time-series in the volumes of *JABA* spanning 1990 (volume 23, issue 1) to 1995 (volume 28, issue 1). Forty-seven percent of the functional analysis data sets (N=235) and 40% of the conventional baselines (N=782) contained less than five observations and were excluded from the present analysis.

A summary of the means of the autocorrelation coefficients computed on the original and detrended data for the 1990-1995 *JABA* sample, its two subsamples, and the Busk and Marascuilo sample appears in Table 4. The frequency distributions associated with these samples appear in Figure 5. Further analysis of the findings follows.

Table 4

Unweighted Means on r_1, r_{1R}, r_{F1}, and r_{HM} Computed for Four Samples

Data Series	Auto-correlation Estimator	1990-1995 *JABA* Sample	Conventional Baseline Subsample	Functional Analysis Subsample	Busk and Marascuilo Sample
Original	r_1	-.020	-.020	-.020	.076
Original	r_{F1}	.119	.108	.151	.219
Detrended	r_{1R}	-.175	-.168	-.194	-.038
Detrended	r_{HM}	.029	.023	.046	.126

Conventional Autocorrelation Estimator r_1 Applied to Original Data

Figure 5 (panel A) shows the frequency distribution of the r_1 autocorrelation estimates computed on the original observations (r_1) from the 1990-1995 *JABA* sample. The unweighted mean (i.e., the arithmetic average) of this distribution is -.020. Panels B and C show the frequency distribution of the r_1 autocorrelation estimates for the conventional baseline and functional analysis subsamples of the 1990-1995 *JABA* sample. The general shapes of these distributions are consistent with the overall pattern of the larger sample. The distribution for the Busk and Marascuilo sample is shown in panel D. This distribution has a higher mean ($\bar{r}_1$ = .076), with a larger proportion of positive than negative values.

r_{F1} Autocorrelation Estimator Applied to Original Data

The autocorrelation estimator r_1 is known to be negatively biased, thus the bias may affect any conclusions based on a meta-analysis employing this estimator. Hence, the reduced bias autocorrelation estimator r_{F1} was applied to all data sets. It can be seen in Figure 6 that the results based on applying r_{F1} to the original data series differ from those based on the r_1 estimator. The r_{F1} distribution in panel A appears to be stretched toward the upper and lower limits and is centered around a higher positive mean (.119) value than the distribution of r_1.. When separated into the conventional baseline (N=150) and functional analysis (N=50) subsamples, the same overall pattern is evident (Figure 6, Panels B and C). These distributions are centered around .108 and .151 respectively, and both are pulled out to the upper and lower limits. Again, the

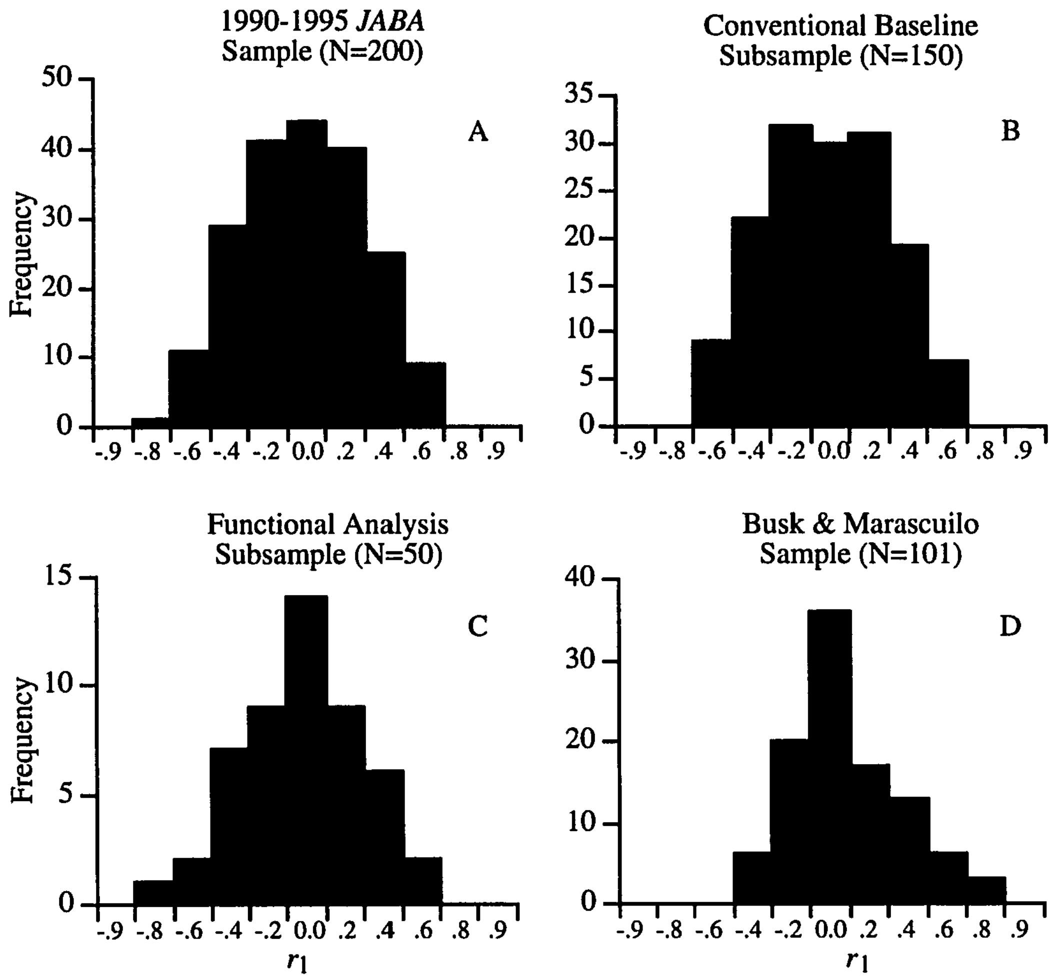

Figure 5. Frequency Distributions for the r_1 Coefficients Computed on Four Samples.

same pattern is clearly evident in the distribution of the r_{F1} values from the Busk and Marascuilo sample (Figure 6, panel D).

t-values on Slopes as Standardized Descriptors of Trend in Original Data

Figure 7 (panel A) shows the frequency distribution of *t* values associated with the slopes (b_1) from individual data sets in the 1990-1995 *JABA* sample (N=200).

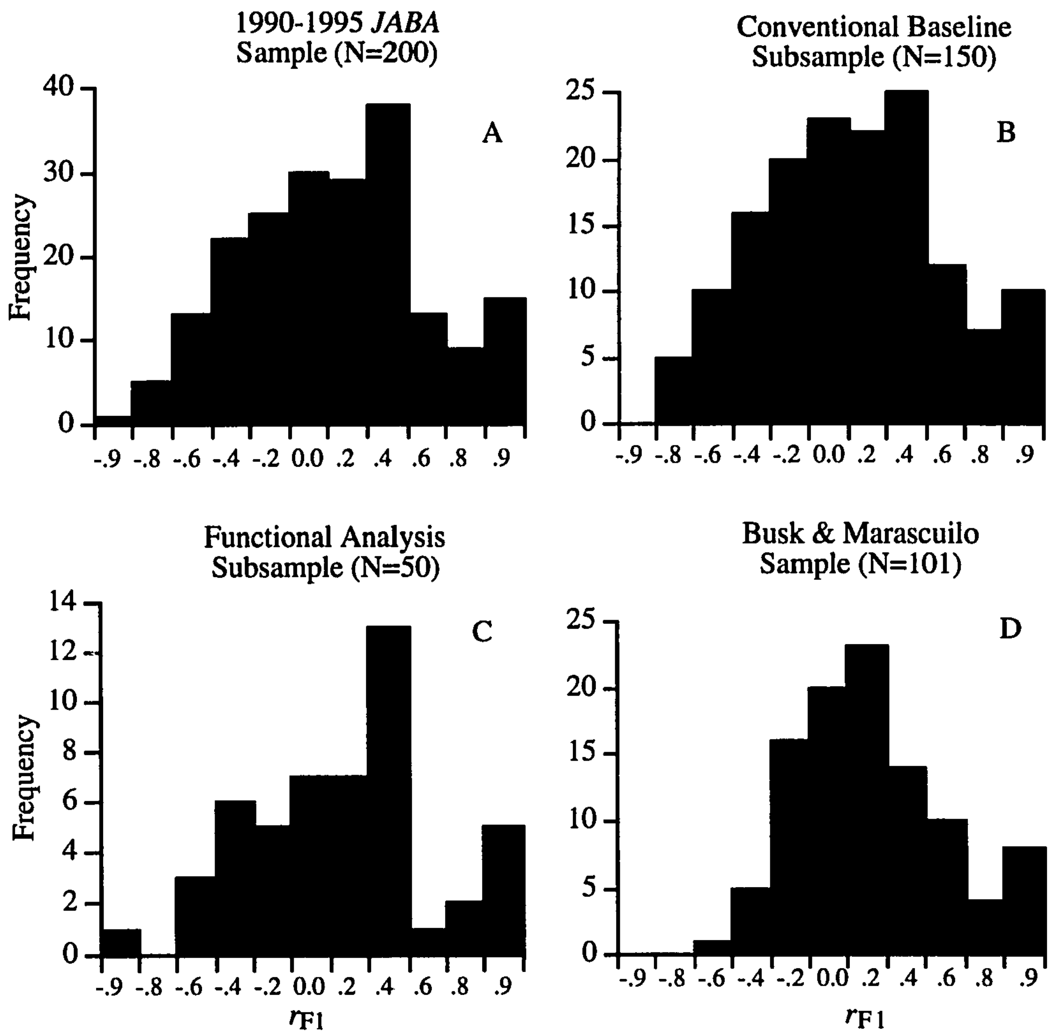

Figure 6. Frequency Distributions for the r_{F1} Coefficients Computed on Four Samples.

Although the distribution is centered close to zero ($\overline{t_{b_1}}$ = -.108), it can be seen that there is a large number of substantial values in the distribution. Values greater than approximately two identify statistically significant trend. Similar patterns are seen when the distributions are plotted for the conventional baseline and functional analysis subsamples and for the Busk and Marascuilo sample (panels B, C, and D of Figure 7). Slope in the data contributes an artificial positive inflation in both the r_1 and r_{F1} values.

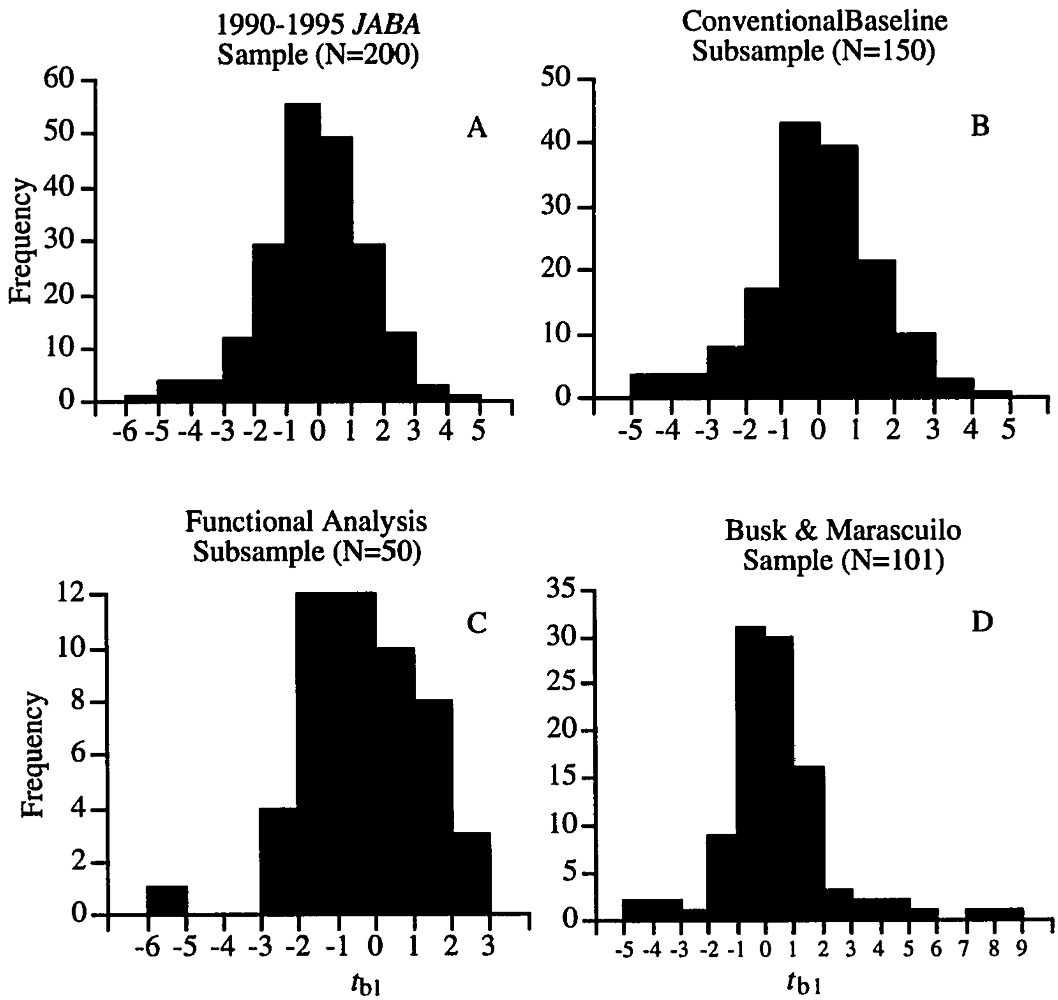

Figure 7. Frequency Distributions for the t_{b1} Coefficients Computed on Four Samples.

Figure 8 contains bivariate distributions of the r_{F1} values plotted on the absolute values of the standardized slopes for the 1990-1995 *JABA* sample and the Busk and Marascuilo sample. These graphs demonstrate the effects of slope on the r_{F1} coefficients. In both cases, higher absolute values for the slopes are strongly associated with higher positive values on r_{F1}.

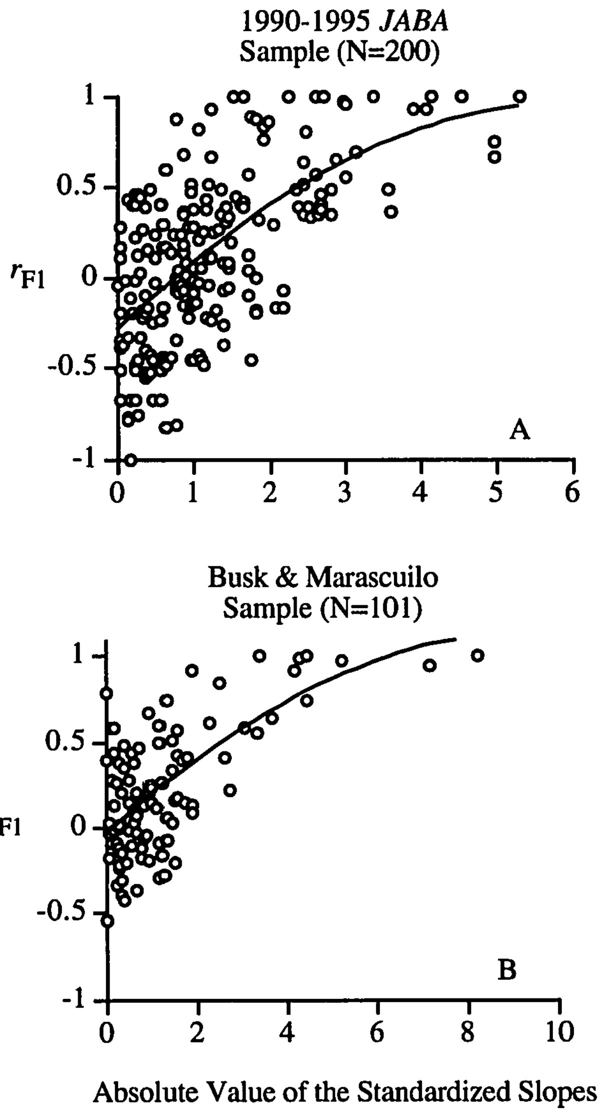

Figure 8. r_{F1} Values Plotted Against the Absolute Values of the Standardized Slopes for the 1990-1995 *JABA* Sample and the Busk and Marascuilo Sample.

The lines fit through the two distributions in Figure 8 are based on quadratic regressions of the r_{F1} values on the absolute standardized slopes and the squared values of those standardized slopes. These results lead to the conclusion that autocorrelation coefficients are strongly affected by trends in the data. The results of linear detrending can be seen in the next section by comparing the autocorrelation based on the original data with those based on detrended data. The influence of nonlinear trend will be discussed subsequently.

<u>Conventional Autocorrelation Estimates Applied to Linear Detrended Data</u>

It can be seen in Figure 9 that when the positive bias resulting from linear trend is removed from the data, the frequency distribution of the r_1 values is centered lower on the scale. Panel A shows the frequency distribution of the r_1 autocorrelation estimates computed on the detrended observations (r_{1R}) contained in the individual data sets of the 1990-1995 *JABA* sample; the mean of this distribution is -.175. Again, this pattern is consistent for the conventional baseline and functional analysis subsamples of the 1995-1990 *JABA* sample (panels B and C of Figure 9, respectively). Panel D shows that the Busk and Marascuilo sample contains a larger proportion of positive than negative r_{1R} coefficients in comparison to panels A through C.

The r_{1R} coefficients, although devoid of the positive bias introduced by linear trend, contain the negative bias inherent in the r_{1R} estimator. It is for this reason that the reduced bias estimator r_{HM} was also applied to the detrended data.

<u>r_{HM} Autocorrelation Estimator Applied to Linear Detrended Data</u>

The r_{HM} autocorrelation coefficient provides a less biased estimate of the degree of autocorrelation in applied behavioral studies than does the conventional estimator r_{1R}. Figure 10 indicates that the r_{HM} coefficients plotted as frequency distributions for

1990-1995 *JABA* sample, the conventional baseline and functional analysis subsamples and the Busk and Marascuilo sample appear more normal than the r_{F1} distributions. Further, these distributions are centered close to zero, and have fewer coefficients in the extreme positive and negative tails. That is, when the effects of both the negative bias inherent in the r_{1R} autocorrelation estimator is reduced and the positive bias introduced by trend in the data is removed, the empirical distributions more closely approximate

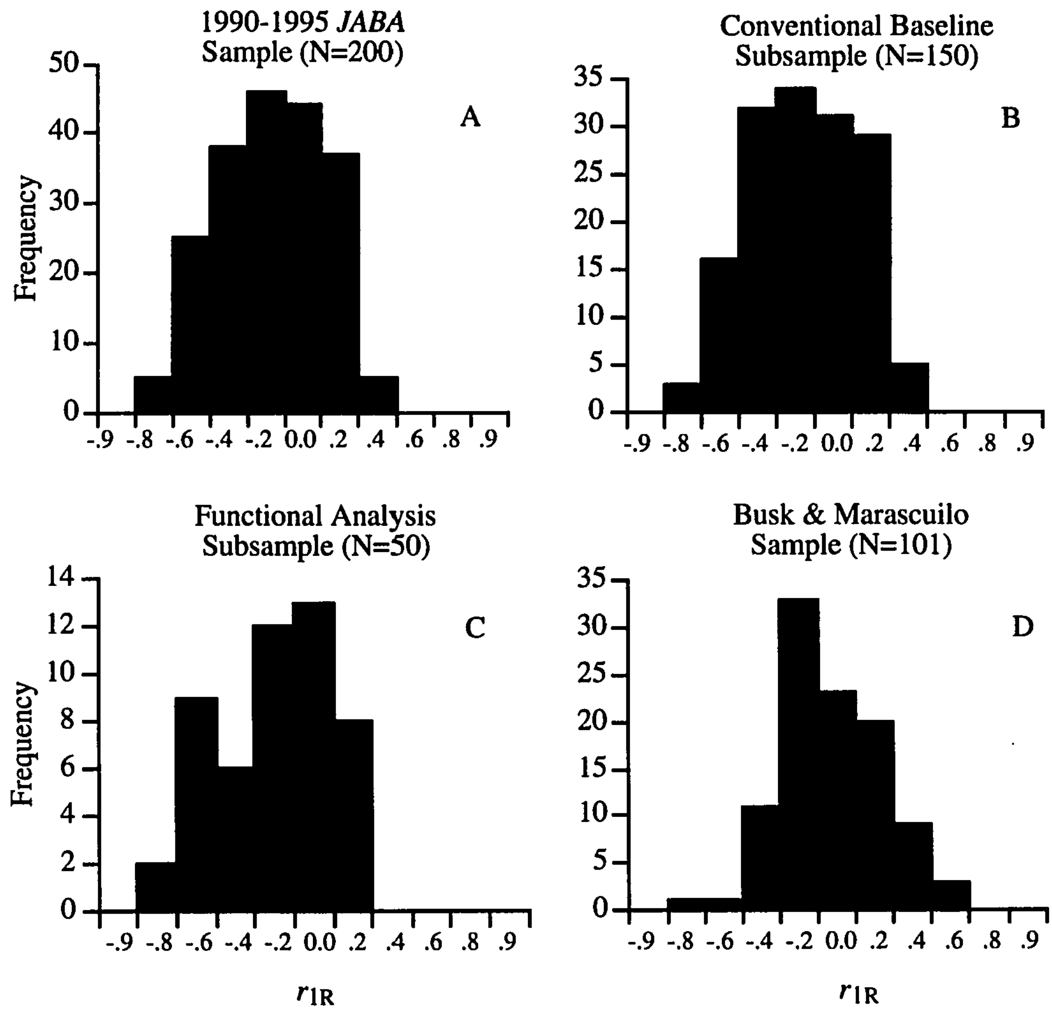

Figure 9. Frequency Distributions for the r_{1R} Coefficients Computed on Four Samples.

normality and are centered around zero.

A summary plot of the unweighted means of the r_1, r_{1R}, r_{F1}, and r_{HM} distributions for the 1990-1995 *JABA* sample, the two subsamples of that sample, and for the Busk and Marascuilo sample can be seen in Figure 11. This figure illustrates the effects of the different estimators and detrending for each sample. In order to assess whether the variation of the r_{F1}, and r_{HM} distributions around zero is more than

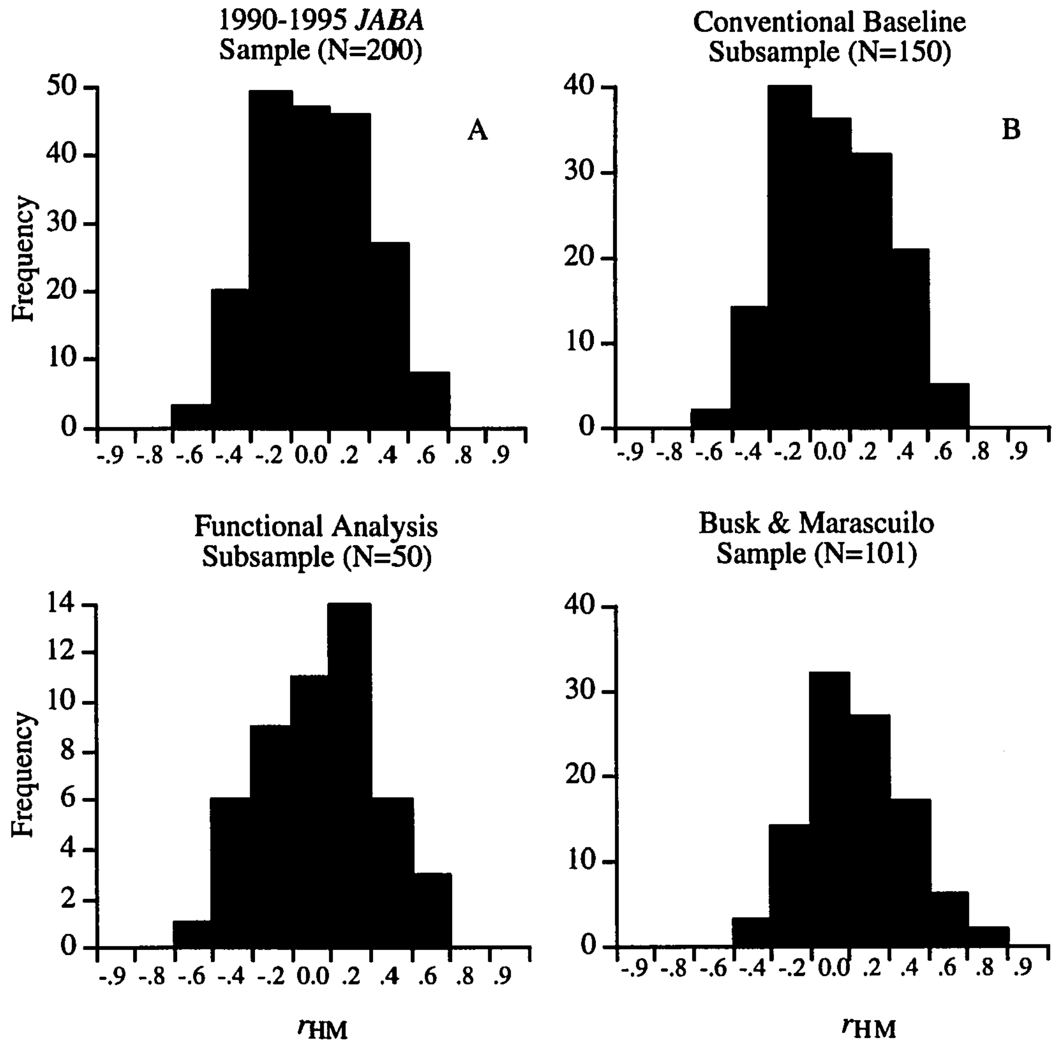

Figure 10. Frequency Distributions for the r_{HM} Coefficients Computed on Four Samples.

that which would be expected on the basis of sampling alone, test statistics associated with the r_{F1} and r_{HM} distributions were computed. The results of these and other inferential tests are described in the subsequent two sections.

Inferential Procedures

Hypothesis tests of H_0: $\rho_1 = 0$ associated with the reduced biased autocorrelation estimators for the original data series and for the linear detrended data were computed. The distribution of the z_{F1} test statistic associated with the r_{F1} distributions indicates that there were more coefficients falling in the tails of the distributions than would be expected if the coefficients differed from zero only as a result of sampling error. However, the distributions of the z_{HM} test statistic associated with the r_{HM} distributions show that the number of coefficients falling in the extreme tails are consistent with that which would be predicted on the basis of sampling error

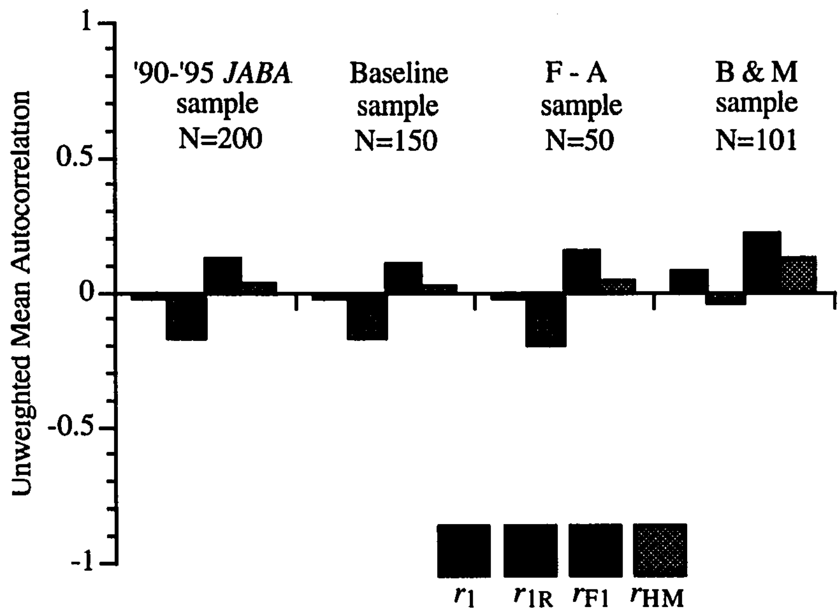

Figure 11. Mean Unweighted Autocorrelation Estimators r_1, r_{1R}, r_{F1}, and r_{HM} for Four Samples.

alone. That is, when the negative bias of the r_1 estimator is controlled for, and the positive bias introduced by linear trend in the data is eliminated, fewer than 5% of the coefficients are more extreme than the critical values of ±2.0. Further description of the inferential analysis based on test statistics follows.

z_{F1} and z_{HM} Distributions

Table 5 contains the means of the test statistics z_{F1} and z_{HM} computed for the r_{F1} and r_{HM} distributions for the 1990-1995 *JABA* sample, its two subsamples, and the Busk and Marascuilo sample. The mean of the z_{F1} distribution for all samples falls between .347 and .701, while the z_{HM} distributions are centered closer to zero. Inspection of the distribution of z_{F1} test statistics (Figure 12) shows that all samples contain more coefficients in the tails than would be predicted on the basis of sampling error alone. The proportion of coefficients falling beyond ±2.0 ranges from 9.33% to 13.86% for the four distributions; 5% or less would be predicted if the coefficients differed from zero only on the basis of sampling error.

Inspection of the z_{HM} distributions (Figure 13) confirms that the large proportion of extreme z_{F1} scores (Figure 11) can be explained by the presence of linear

Table 5

Mean Test Statistics Computed on the r_1, $r_{1\mathrm{R}}$, r_{F1}, and r_{HM} Distributions for the Four Samples

Test Statistic	1990-1995 *JABA* Sample	Conventional Baseline Subsample	Functional Analysis Subsample	Busk and Marascuilo Sample
z_{F1}	.352	.347	.368	.701
z_{HM}	.149	.139	.176	.567

trend in the individual data sets. That is, the distributions of test statistics associated with the reduced bias coefficients computed on the detrended data (r_{HM}) contain fewer than 5% more extreme than ±2.0. The actual numbers of coefficients found beyond ±2.0 and their corresponding percentages for the z_{F1} and z_{HM} distributions can be found in Table 6.

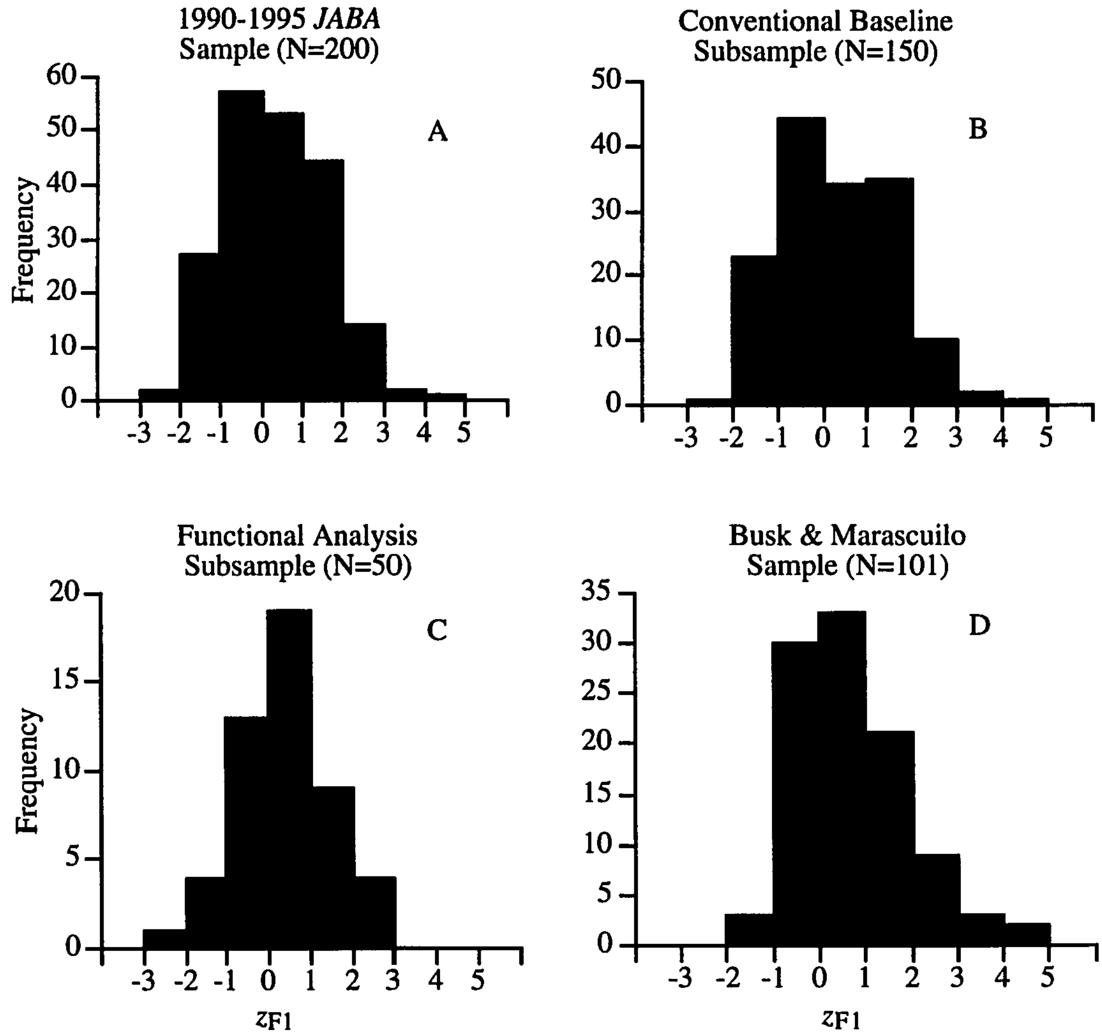

Figure 12. Frequency Distributions for the z_{F1} Coefficients Computed on Four Samples.

Chi-square Tests of the Homogeneity of r_{F1} and r_{HM} Autocorrelation Coefficients

The chi-square statistics used to test the homogeneity of the obtained r_{F1} and r_{HM} autocorrelation distributions around their central (weighted mean) values can be found in Table 7. The obtained χ^2 values associated with the r_{F1} distributions for the 1990-1995 *JABA* sample, for the conventional baseline subsample, and for the Busk and Marascuilo sample were greater than their respective critical values ($_{.05}\chi^2_{J-1}$). These

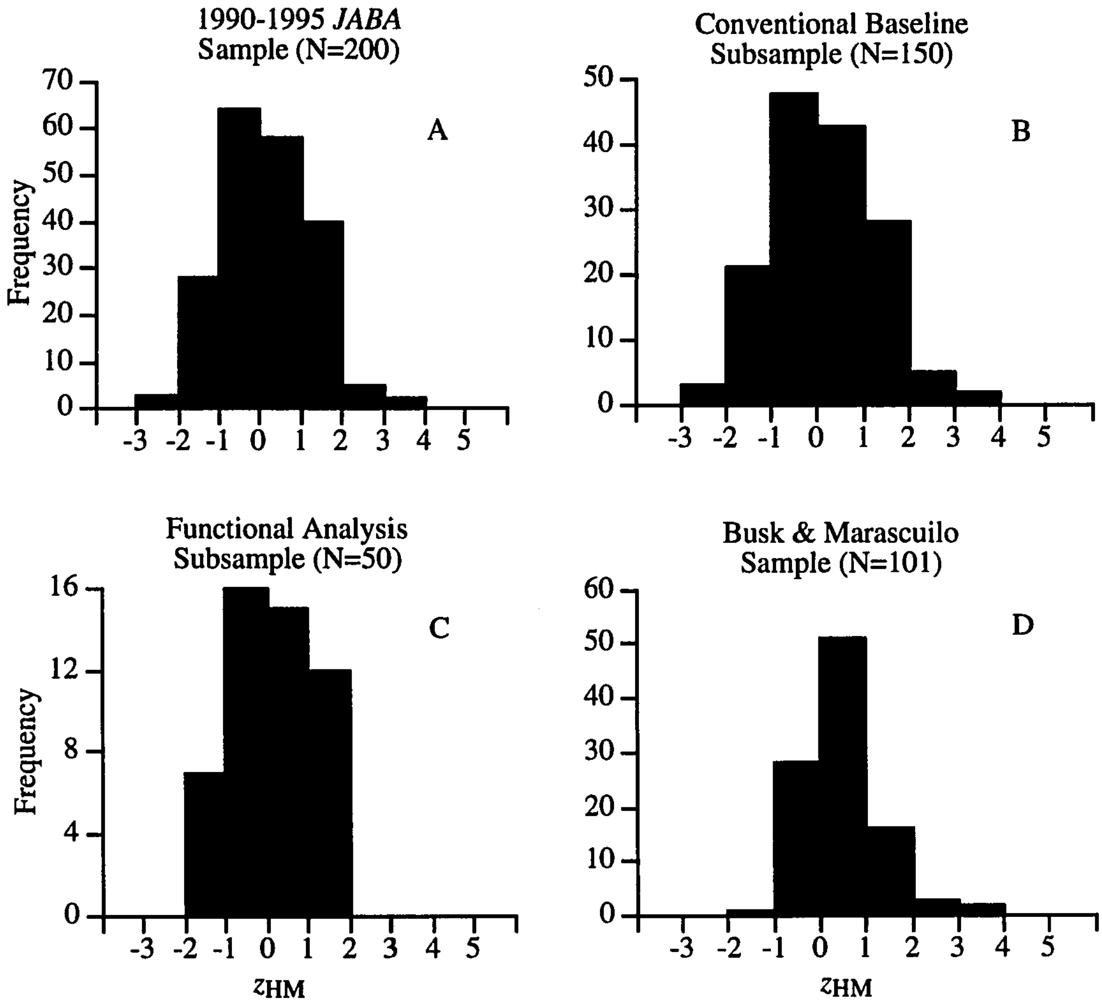

Figure 13. Frequency Distributions for the z_{HM} Coefficients Computed on Four Samples.

Table 6

Number of Test Statistics Falling Beyond the ±2.0 Limits for z_{F1} and z_{HM} Distributions for the 1990-1995 *JABA* Sample, the Conventional Baseline and Functional Analysis Subsamples, and the Busk and Marascuilo Sample

Test Statistic	1990-1995 *JABA* Sample	Conventional Baseline Subsample	Functional Analysis Subsample	Busk and Marascuilo Sample
z_{F1}	19 (9.5%)	14 (9.33%)	5 (10.0%)	14 (13.86%)
z_{HM}	6 (3.0%)	6 (4.0%)	0 (0.0%)	5 (4.95%)

results indicate that the ρ_j parameters associated with the process in the behavioral studies are not homogenous. The obtained χ^2 value associated with r_{F1} for the functional analysis subsample was less than the critical value of ${}_{.05}\chi^2_{49}$, thus the data are consistent with the hypothesis of a common ρ and it is appropriate to pool these coefficients for significance testing against H_0: $\rho = 0$.

The results of the tests associated with the r_{HM} distributions show that when the influence of trend is removed from the autocorrelation estimates, the 1990-1995 *JABA* sample and the Busk and Marascuilo sample, as well as the functional analysis subsample are associated with nonsignificant chi-square values (p values range from .06 to .58). These distributions of r_{HM} values are considered homogenous; there is justification for pooling the coefficients to obtain the pooled mean r_{HM} coefficient upon which to construct 95% confidence intervals on ρ. The r_{HM} distribution for the conventional baseline subsample yields a significant chi-square statistic (p=.03); the distribution appears to contain heterogeneous coefficients. Hence there is limited interest in pooling the coefficients because there is not a common process autocorrelation across the various studies in the sample.

Table 7

Chi-Square Obtained Values Associated With r_{F1} and r_{HM} for Four Samples

χ^2_{J-1}	1990-1995 *JABA* Sample	Conventional Baseline Subsample	Functional Analysis Subsample	Busk and Marascuilo Sample
χ^2_{J-1} Obtained$_{F1}$	282.53* (p=.00)	232.77* (p=.00)	55.62 (p=.24)	142.75* (p=.00)
χ^2_{J-1} Obtained$_{HM}$	230.40 (p=.06)	184.14* (p=.03)	46.27 (p=.58)	114.94 (p=.15)

Confidence Intervals on ρ_1

Table 8 contains the pooled weighted r_{F1} and r_{HM} autocorrelations computed for the four samples. The coefficients associated with the original data series for each sample are considerably larger than those associated with the detrended data. Though it appears that there is little difference between the conventional baseline and functional analysis subsamples of the 1990-1995 *JABA* sample and that the weighted coefficient computed for the Busk and Marascuilo sample is much larger, examination of the 95% confidence intervals on ρ_1 provide more adequate descriptions of the process autocorrelations.

Table 9 contains the confidence intervals on the process autocorrelations for the four samples based on the pooled weighted mean r_{HM}. As expected, the only interval containing zero is that of the functional analysis subsample. The statistical model that explains the obtained functional analysis autocorrelation coefficients best is the simple model $r_{\mathrm{HMj}} = \varepsilon_j$, where the observed r_{HM} is a function of sampling errors ε_j that have a mean of zero.

The remaining three confidence intervals (i.e., those of the 1990-1995 JABA sample, the conventional baseline subsample, and the Busk and Marascuilo sample) do not contain zero, although they are relatively narrow. The statistical model that explains the obtained autocorrelation coefficients for these samples is the three-component model $r_{\mathrm{HMj}} = \bar{\rho}_1 + \left(\rho_{1j} - \bar{\rho}_1\right) + \varepsilon_j$, where the observed $\bar{r}_{\mathrm{HM}}$ is a function of the overall mean of the process autocorrelation coefficients ($\bar{\rho}$), the difference between ρ_j and $\bar{\rho}$, and sampling error.

Table 8

Weighted Means on r_{F1} and r_{HM} Computed for Four Samples

Autocorrelation Estimator	1990-1995 *JABA* Sample	Conventional Baseline Subsample	Functional Analysis Subsample	Busk and Marascuilo Sample
r_{F1}	.144	.123	.156	.218
r_{HM}	.056	.057	.053	.139

Table 9

95% Confidence Intervals Around $\bar{r}_{\mathrm{HM}}$ on the Process Autocorrelations Computed for Four Samples

1990-1995 *JABA* Sample	Conventional Baseline Subsample	Functional Analysis Subsample	Busk and Marascuilo Sample
(.02, .09)	(.02, .10)	(-.04, .14)	(.09, .19)

The results of the confidence intervals around the mean weighted coefficients suggest that the mean process autocorrelation is close to zero after removing the effects of linear trend. A second issue, beyond the mean of the distributions, is whether the variance of the distributions is explainable as sampling error. To appropriately determine the amount of observed variance in r_{HM} values that can be attributed to sampling error relative to the amount that can be attributed to variation in process autocorrelation, estimates of both contributions were computed. These results are presented next.

Estimates of Error Variance, Variance in the r_{HM} Distributions, and Variance Due to Other Sources

The homogeneity test on the r_{HM} coefficients reported above indicate that a reasonable statistical model for the variance in the r_{HM} distribution for the functional analysis subsample is:

$$\sigma^2_{r_{HM}} = \sigma^2_{\varepsilon}.$$

This model suggests that the variation in the obtained distribution is solely due to sampling error. A reasonable model for the r_{HM} variance in the 1990-1995 *JABA* sample, the conventional baseline subsample and the Busk and Marascuilo sample includes the error term, plus an additional term to explain differences in process autocorrelations, and can be written as:

$$\sigma^2_{r_{HM}} = \sigma^2_{\rho_1} + \sigma^2_{\varepsilon}.$$

If the variance of the r_{HM} values is a function of (a) error variance and (b) variance in process autocorrelations, then the amount of variance observed above that predicted on the basis of sampling error can be interpreted as the variance of the unknown ρ_{1j} values. The latter variance can be estimated from the sample data and knowledge of the theoretical variance of the sample r_{HM} values. Table 10 shows the variance estimates associated with the r_{F1} and the r_{HM} distributions computed for the four samples. Rows 1 and 2 contain the estimates for the processes containing linear trend. Row 3 contains the differences between rows 1 and 2 ($\hat{\sigma}^2_{r_{F1}}$ - $\hat{\sigma}^2_{\varepsilon}$); these differences estimate the amount of variation $\hat{\sigma}^2_{\rho_j}$ in the distributions beyond that explained by sampling error. Because the processes upon which the sample r_{F1} autocorrelations are based are known to contain linear trend, it is not surprising to find that all computed values of $\hat{\sigma}^2_{\rho_j}$ are larger than zero, although the estimate for the functional analysis is substantially smaller than for the other samples.

Rows 4 and 5 of Table 10 contain the variance estimates for the linear detrended distributions. The difference between rows 4 and 5 ($\hat{\sigma}^2_{r_{HM}}$ - $\hat{\sigma}^2_{\varepsilon}$) is the amount of variation remaining in the distributions above that predicted by sampling error alone ($\hat{\sigma}^2_{\rho_j}$). The values of $\hat{\sigma}^2_{\rho_j}$ computed on the three samples with corresponding significant chi-square test statistics (i.e., the 1990-1995 *JABA* sample, the conventional baseline subsample of that sample, and the Busk and Marascuilo sample) range from .0090 to.0137. The $\hat{\sigma}^2_{\rho_j}$ value for the functional analysis subsample is a small negative (-.007) which indicates that sampling error accounts for all of the variation in the r_{HM} distribution for this sample. The proportion of the estimated variance of the r_{F1} and r_{HM} distributions that appears to be explained by variance among the process ρ_j values can be seen in Table 11.

Table 10

Variance Estimates $\hat{\sigma}_r^2$ and $\hat{\sigma}_\varepsilon^2$ Computed for Four Sample Distributions

Row	Auto-correlation Estimator	Variance Estimates	1990-1995 *JABA* Sample	Conventional Baseline Subsample	Functional Analysis Subsample	Busk and Marascuilo Sample
1)	r_{F1}	$\hat{\sigma}^2_{r_{F1}}$	.158	.153	.190	.127
2)		$\hat{\sigma}^2_\varepsilon$	.115	.102	.181	.094
3)		$\hat{\sigma}^2_{r_{F1}} - \hat{\sigma}^2_\varepsilon$	.043	.051	.009	.033
4)	r_{HM}	$\hat{\sigma}^2_{r_{HM}}$	.062	.061	.065	.058
5)		$\hat{\sigma}^2_\varepsilon$	.052	.048	.072	.049
6)		$\hat{\sigma}^2_{r_{HM}} - \hat{\sigma}^2_\varepsilon$	.010	.014	-.007	.009

Table 11

Estimated Proportion of Total Variance in r_{F1} and r_{HM} Explained by Process Autocorrelation Variance ($\hat{\sigma}^2_{\rho_1}$) and Error Variance ($\hat{\sigma}^2_\varepsilon$)

Series	AC Estim-ator	Proportion explained by	1990-1995 *JABA* Sample	Conventional Baseline Subsample	Functional Analysis Subsample	Busk and Marascuilo Sample
Original	r_{F1}	$\hat{\sigma}^2_{\rho_1}$	.272	.333	.047	.26
		$\hat{\sigma}^2_\varepsilon$	.728	.667	.953	.74
De-trended	r_{HM}	$\hat{\sigma}^2_{\rho_1}$	.161	.23	.00	.16
		$\hat{\sigma}^2_\varepsilon$	.839	.77	1.0	.84

DISCUSSION

Many empirical and expository articles on the "autocorrelation problem" with single-subject time-series data have appeared in the behavioral literature (Busk & Marascuilo, 1988; Gottman & Glass, 1978; Sharpley & Alavosius, 1988; Suen, 1987; Suen & Ary, 1987). These articles generally conclude that ARIMA (and other) complex time-series methods are essential for the analysis of applied behavior analytic data. They argue that the errors of such data sets are autocorrelated. However, previous attempts to characterize the distribution of autocorrelation coefficients associated with a population of applied behavior analytic data have been flawed.

Huitema (1990) presented the most important errors in earlier attempts to characterize autocorrelation in applied behavioral time-series data sets. Early studies ignored intervention effects in computing autocorrelation coefficients, thus using an inappropriate model upon which to base claims of high autocorrelation. This problem was recognized by later researchers, but those researchers failed to go beyond an ANOVA model in explaining applied behavioral data, thus ignoring the effects of trend on the magnitude of autocorrelation estimators (Busk & Marascuilo, 1988; Sharpley & Alavosius, 1988; Suen & Ary, 1987). Finally, the universal estimation problem due to the use of the negatively biased r_1 estimator was recognized in some studies of the "autocorrelation problem," however, no reduced biased estimator was available (Huitema, 1988). The major purpose of the present research was to provide an unbiased estimate of the level of autocorrelation present in the errors of the population of recently published applied behavioral studies through specification of more reasonable statistical models and use of reduced bias estimators of lag-1 autocorrelation.

Conclusions

Efficacy of Reduced Bias Estimators and Linear Detrending

The first analysis conducted in this study involved computing conventional r_1 autocorrelation coefficients on the individual time-series data sets contained in each sample. These computations produced low mean values (-.02 to .08). The negative bias inherent in the conventional r_1 estimator is the source of what Huitema (1990) referred to as the universal estimation problem. Application of the reduced bias r_{F1} estimator produced a collection of higher positive mean coefficients (.11 to .22). These results appear to be consistent with those of Busk and Marascuilo (1988). That is, when the unbiased estimator (r_{F1}) is used, there appears to be a significant proportion of behavioral data sets exhibiting autocorrelation, generally in the positive direction. Most of the apparent positive autocorrelation can, however, be explained by linear trend in the data. This characterizes what Huitema (1990) labeled model misspecification type II.

Application of the conventional r_1 estimator on linear detrended data produced negative mean autocorrelation coefficients ranging from -.04 to -.12. Computation of these conventional autocorrelation estimators on detrended data corrected for the model misspecification flaw, but the results retained the negative bias of r_1. For this reason, a reduced bias estimator r_{HM} was employed. These results showed that when the data are appropriately detrended and a reduced bias estimator is employed, the autocorrelation distribution is similar to that expected by chance when sampling from a population of independent (nonautocorrelated) errors.

The conclusion that behavioral data are not highly autocorrelated is drawn. In most cases the explanation for "apparent" autocorrelation is not that the errors are autocorrelated. Rather, the statistical model used to explain the data has been misspecified. Artificial autocorrelation is to be expected if trend (linear or otherwise) and other deterministic components in the data have not been specified in the statistical model.

Potential Moderating Variables

A second issue, beyond the mean of the distributions was addressed in this study; that is, "What proportion of the variance of the r_{HM} distributions is explainable as sampling error?" It was concluded that sampling error alone accounts for the total variation in obtained r_{HM} values for the functional analysis subsample. A portion of variation beyond sampling error is left to be explained for the remaining three samples. Several sources of artificial autocorrelation were postulated that included model misspecification problems. These problems were apparent when inspecting the original data plots from a collection of sampled studies. Other potential moderator variables that may account for the variation beyond sampling error are suggested by the differing results found for the functional analysis subsample when compared to the three other samples. The samples employed in this study suggest that "baseline type" can function as a moderator variable influencing the magnitude of autocorrelation in the 1990-1995 *JABA* sample.

Functional analysis data sets are a more homogeneous grouping than are the conventional baseline time series. In functional analyses the analyst identifies environmental variables that may influence the level of target behavior. These variables are precisely controlled by the investigator and baseline rates of behavior are obtained under the different conditions. The behavior of interest is typically self injury, but other topographies such as stereotypy may also be assessed under this assay. Thus, there are several dimensions upon which functional analysis and conventional baselines differ. The functional analysis assay is used to study a narrow range of problematic behavioral topographies, the metrics typically employed are measures of rates or proportions of behavior, sessions are generally short (10-15 minutes), the characteristic subject is a person with a developmental or related disability, and data are gathered under well controlled experimental conditions. The collection of conventional baselines that make up the remaining portion of the sampled *JABA* studies vary on every dimension. A wide range of behavioral topographies are studied using an equally wide range of metrics. Sessions vary in length from short samples (e.g., 5-10 minutes) to large chunks of time (e.g., the work day, school day, or week), there is no one characteristic subject, and data are gathered under conditions ranging from tight experimental control to community settings where control is nearly impossible. Any one of these dimensions may prove to be a moderator variable that contributes to autocorrelation of errors in applied behavioral data.

Suggestions for Future Research

It is suggested that future research would benefit from two methodological changes. First, the reduced bias r_{HM} autocorrelation estimator for use with detrended data is not completely free of bias. At low values of ρ, r_{HM} will produce essentially unbiased estimates of the process autocorrelation. When ρ is a large value, r_{HM} is biased toward zero. Although the bias is less than that of r_1, the results of employing r_{HM} with a process autocorrelation of large magnitude is to pull coefficients with extreme values in toward the center of the distribution. A reduced biased estimator for use on detrended data corresponding to the r_{F1} coefficient is currently under development (McKean, McKnight & Huitema, in preparation). Future research on autocorrelation in applied behavioral data could benefit by employing an autocorrelation estimator that is less biased at extreme values of ρ_1.

A second direction for future research involves modeling the complete design. Only baseline phases with a minimum of six observations were sampled in this study. Applied behavioral studies typically contain several baseline, intervention, and follow-up phases, often with fewer than six data points in a single phase. Thus, the sampled phases do not wholly reflect the characteristics of applied behavioral experimental designs. Future research employing reduced bias estimators should include attempts to model pre- and post-intervention phases across a variety of single-subject design types.

Summary Recommendations for Analysts

This study has shown that autocorrelation of the errors of a reasonable model is not a problem for applied behavioral time-series data. Complex time-series analysis procedures such as ARIMA models are not necessary, and are generally inappropriate, for analyzing applied behavioral data. Alternative intervention models such as ANOVA and interrupted time-series regression models generally provide a reasonable fit to behavioral data. When autocorrelated errors are present a new procedure based on generalized least squares is now available (McKnight, McKean, & Huitema, in preparation).

The sound conclusions drawn by Huitema (1988) stand as appropriate recommendations for the analysis of behavior analytic time-series data. His conclusions are repeated as follows:

1. The residuals of appropriate conventional linear models fit to a population of typical applied behavioral studies have little or no autocorrelation.

2. The major implication of little or no autocorrelation is that conventional statistical methods based on the general linear model should be considered for the statistical analysis of behavioral data.

3. The most appropriate method for a given study will depend on the data and the researcher's question. In many cases Shewhart charts, one-factor analysis of variance, or piecewise-discontinuity regression models will provide a satisfactory solution.

4. The use of time-series models should be considered if sufficient data are available to justify the use of such models.

However, it should be understood that though these models generally fit, there are assumptions other than independence that should be satisfied before adopting any general linear model. Examples of data that cannot be fit with such models, or with any conventional time-series model, are routinely encountered in the behavioral literature. For example, Figures 14, 15 and 16 present data indicating obvious departures from (a) the constant variance assumption, and (b) the normality assumption. The use of conventional models in the situations described in these figures are very likely to lead to inappropriate conclusions. The use of statistical models that are inconsistent with the data should be discouraged.

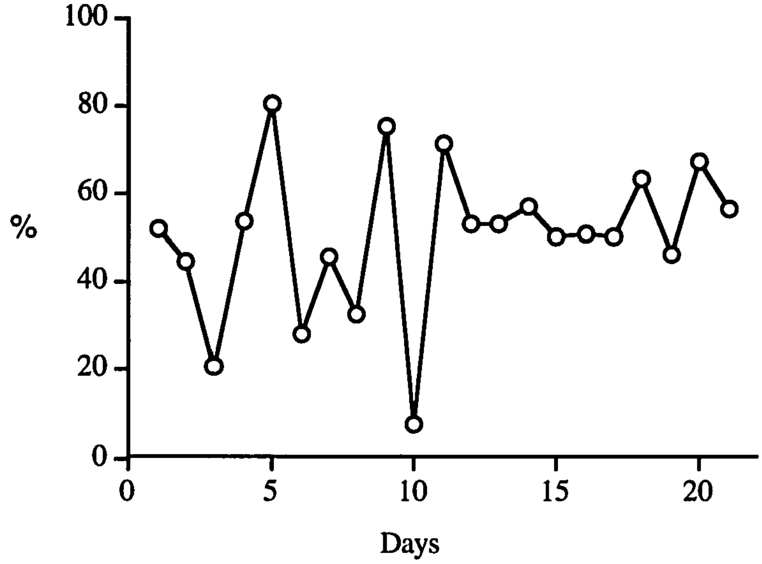

Figure 14. Example of Nonconstant Variance in a Conventional Baseline Phase (Adapted from Austin, Hatfield, Grindle, and Bailey, 1993).

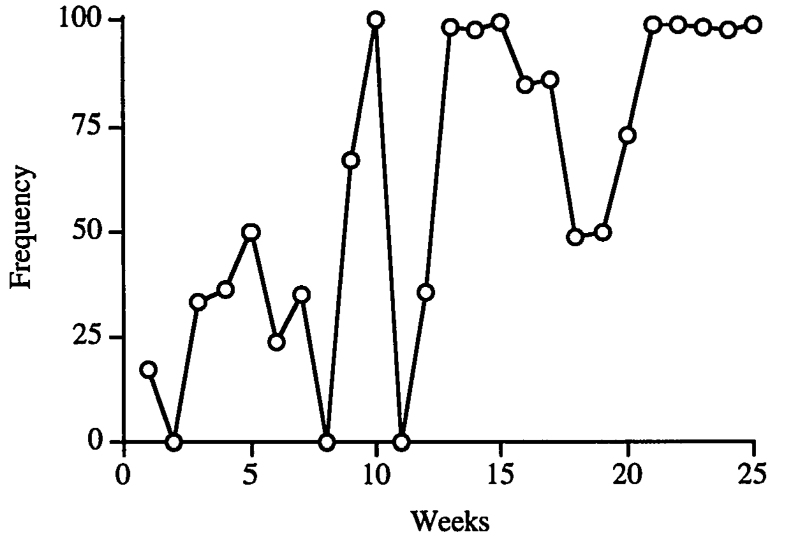

Figure 15. Example of Extreme Instability in a Conventional Baseline Phase (Adapted from Charlop and Trasowech, 1991;).

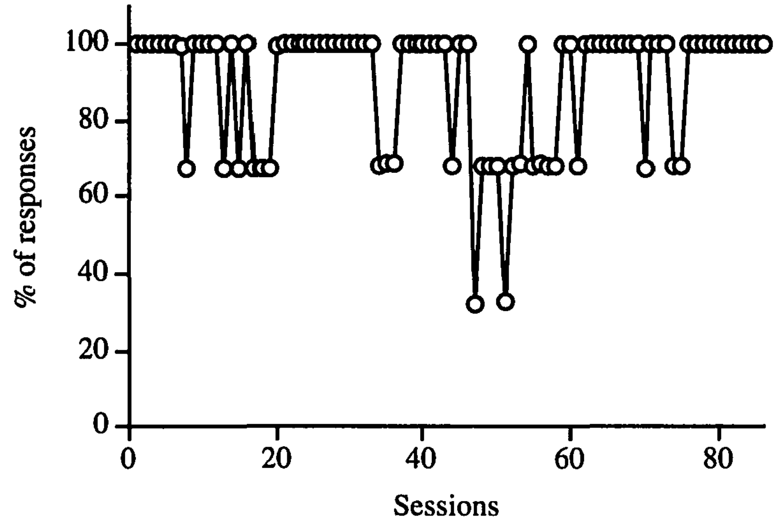

Figure 16. Example of Ceiling Effect in a Conventional Baseline Phase (Adapted from Krantz, McClannahan, and Poulson, 1994;).

Appendix A

List of Figures From Which the Conventional Baseline Subsample (N=150) of the 1990-1995 *JABA* Sample Was Drawn

For ease of retrieval, the following list of baseline figures (N=150) and the articles that they were drawn from is organized according to the order in which they appear in the *Journal of Applied Behavior Analysis* beginning with Issue 1, Volume 23 (1990). Each article listed generally comprises more than one data set, and two or more data sets may have been drawn from the same article or figure, therefore each sample is identified with sufficient information to clearly distinguish it from other data sets within the figure or article.

Iwata, B. A., Pace, G. M., Kalsher, M. J., Cowdery, G. E., & Cataldo, M. F. (1990). Experimental analysis and extinction of self-injurious escape behavior. Journal of Applied Behavior Analysis, 23, 11-28.
Figure 2, Noreen, SIB, page 20
Figure 2, Noreen, compliance, page 20

France, K. G., & Hudson, S. M. (1990). Behavior management of infant sleep disturbance. Journal of Applied Behavior Analysis, 23, 91-98.
Figure 1, child 7, page 95
Figure 2, child 2, page 96

Burgio, L. D., Engel, B. T., Hawkins, A., McCormick, K., Scheve, A., & Jones, L. T. (1990). A staff management system for maintaining improvements in continence with elderly nursing home residents. Journal of Applied Behavior Analysis, 23, 111-118.
Figure 1, Jenny, page 116

Koegel, R. L., & Koegel, L. K. (1990). Extended reductions in stereotypic behavior of students with autism through a self-management treatment package. Journal of Applied Behavior Analysis, 23, 119-128.
Figure 1, student 3, page 122
Figure 2, student 3, page 124

Ward, W. D., & Stare, S. W. (1990). The role of subject verbalization in generalized correspondence. Journal of Applied Behavior Analysis, 23, 129-137.
Figure 2, Jen, free-choice table, page 134

Alavosius, M. P., & Sulzer-Azaroff, B. (1990). Acquisition and maintenance of health-care routines as a function of feedback density. Journal of Applied Behavior Analysis, 23, 151-162.
Figure 1, top panel, page 156
Figure 1, fifth panel, page 156

Charlop, M. H., Kurtz, P. F., & Casey, F. G. (1990). Using aberrant behaviors as reinforcers for autistic children. Journal of Applied Behavior Analysis, 23, 163-182.
Figure 7, bottom panel, page 177

Parsons, M. B., & Reid, D. H. (1990). Assessing food preferences among persons with profound mental retardation: Providing opportunities to make choices. Journal of Applied Behavior Analysis, 23, 183-196.

Figure 2, Todd, toast with jelly, page 188
Figure 2, Todd, water, page 188

Matson, J. L., Sevin, J. A., Fridley, D., & Love, S. R. (1990). Increasing spontaneous language in three autistic children. Journal of Applied Behavior Analysis, 23, 227-234.
Figure 2, top panel, page 231

Haring, T. G., & Kennedy, C. H. (1990). Contextual control of problem behavior in students with severe disabilities. Journal of Applied Behavior Analysis, 23, 235-244.
Figure 3, top panel, page 241

Cooper, L. J., Wacker, D. P., Sasso, G. M., Reimers, T. M., & Donn, L. K. (1990). Using parents as therapists to evaluate appropriate behavior of their children: Application to a tertiary diagnostic clinic. Journal of Applied Behavior Analysis, 23, 285-296.
Figure 3, Bob, best condition, page 293
Figure 3, Matt, worst condition, page 293
Figure 3, John, worst condition, page 293

Kohler, F. W., & Greenwood, C. R. (1990). Effects of collateral peer supportive behaviors within the classwide peer tutoring program. Journal of Applied Behavior Analysis, 23, 307-322.
Figure 8, middle panel, page 319
Figure 8, bottom panel, page 319

White, A. G., & Bailey, J. S. (1990). Reducing disruptive behaviors of elementary physical education students with Sit and Watch. Journal of Applied Behavior Analysis, 23, 353-360.
Figure 1, bottom panel, page 357

Osborne, K., Rudrud, E., & Zezoney, F. (1990). Improved curveball hitting through the enhancement of visual cues. Journal of Applied Behavior Analysis, 23, 371-378.
Figure 1, bottom panel, page 374

Love, S. R., Matson, J. L., & West, D. (1990). Mothers as effective therapists for autistic children's phobias. Journal of Applied Behavior Analysis, 23, 379-386.
Figure 1, bottom panel, page 383

Steege, M. W., Wacker, D. P., Cigrand, K. C., Berg, W. K., Novak, C. G., Reamers, T. M., Sasso, G. M., & DeRaad, A. (1990). Use of negative reinforcement in the treatment of self-injurious behavior. Journal of Applied Behavior Analysis, 23, 459-468.
Figure 4, bottom panel, page 464

McClannahan, L. E., McGee, G. G., MacDuff, G. S., & Krantz, P. J. (1990). Assessing and improving child care: A personal appearance index for children with autism. Journal of Applied Behavior Analysis, 23, 469-482.
Figure 1, institution, page 475

Figure 2, fourth panel, page 476

Ludwig, T. D., & Geller, E. S. (1991). Improving the driving practices of pizza deliverers: Response generalization and moderating effects of driving history. Journal of Applied Behavior Analysis, 24, 31-44.
Figure 1, middle panel, page 36

Bell, K. E., Young, K. R., Salzberg, C. L., & West, R. P. (1991). High school driver education using peer tutors, direct instruction, and precision teaching. Journal of Applied Behavior Analysis, 24, 45-52.
Figure 3, learner 3, passing errors, page 49

Ragnarsson R. S., & Bjorgvinsson, T. (1991). Effects of public posting on driving speed in Icelandic traffic. Journal of Applied Behavior Analysis, 24, 53-58.
Figure 2, >70 km/hr, page 57

Piazza, C. C., & Fisher, W. (1991). A faded bedtime with response cost protocol for treatment of multiple sleep problems in children. Journal of Applied Behavior Analysis, 24, 129-140.
Figure 1, bottom panel, page 134.

Sosland, J. M., & Christophersen, E. R. (1991). Does SleepTight work? A behavioral analysis of the effectiveness of SleepTight for the management of infant colic. Journal of Applied Behavior Analysis, 24, 161-166.
Figure 1, second panel, page 166
Figure 1, fourth panel, page 166

Durand, M. V., & Carr, E. G. (1991). Functional communication training to reduce challenging behavior: Maintenance and application in new settings. Journal of Applied Behavior Analysis, 24, 251-264.
Figure 2, middle panel, page 260
Figure 2, bottom panel, page 260

Homer, R. H., Day, H. M., Sprague, J. R., O'Brien, M., & Heathfield, L. T. (1991). Interspersed requests: A nonaversive procedure for reducing aggression and self-injury during instruction. Journal of Applied Behavior Analysis, 24, 265-278.
Figure 3, bottom panel, page 275

Dattilo, J., & Camarata, S. (1991). Facilitating conversation through self-initiated augmentative communication treatment. Journal of Applied Behavior Analysis, 24, 369-378.
Figure 1, bottom panel, page 375

Duker, P. C., & van Lent, C. (1991). Inducing variability in communicative gestures used by severely retarded individuals. Journal of Applied Behavior Analysis, 24, 379-386.
Figure 1, Rosita, page 384
Figure 1, Yashenka, page 384

Green, C. W., Reid, D. H., Perkins, L., & Gardner, S. M. (1991). Increasing habilitative services for persons with profound handicaps: An application of structural analysis to staff management. Journal of Applied Behavior Analysis, 24, 459-472.
Figure 2, middle panel, training, page 467

Kirby, K. C., Fowler, S. A., & Baer, D. M. (1991). Reactivity in self-recording: Obtrusiveness of recording procedure and peer comments. Journal of Applied Behavior Analysis, 24, 487-498.
Figure 1, third panel, page 491

Green, C. W., Reid, D. H., Canipe, V. S., & Gardner, S. M. (1991). A comprehensive evaluation of reinforcer identification processes for persons with profound multiple handicaps. Journal of Applied Behavior Analysis, 24, 537-552.
Figure 2, top panel, page 544

Derrickson, J. G., Neef, N. A., & Parrish, J. M. (1991). Teaching self-administration of suctioning to children with tracheostomies. Journal of Applied Behavior Analysis, 24, 563-570.
Figure 1, bottom panel, doll, page 568

Ingenmey, R., & Van Houten, R. (1991). Using time delay to promote spontaneous speech in an autistic child. Journal of Applied Behavior Analysis, 24, 591-596.
Figure 1, third panel, page 594

Johnston, J. M., Greene, K. S., Rawal, A., Vazin, T., & Winston, M. (1991). Effects of caloric level on ruminating. Journal of Applied Behavior Analysis, 24, 597-603.
Figure 1, Doris, lunch, page 601

Lombard, D., Neubauer, T. E., Canfield, D., & Winett, R. A. (1991). Behavioral community intervention to reduce the risk of skin cancer. Journal of Applied Behavior Analysis, 24, 677-686.
Figure 4, pool B, children, page 684

Cope, J. G., & Allred, L. J. (1991). Community intervention to deter illegal parking in spaces reserved for the physically disabled. Journal of Applied Behavior Analysis, 24, 687-694.
Figure 2, page 690

DeVries, J. E., Burnette, M. M., & Redmon, W. K. (1991). AIDS prevention: Improving nurses' compliance with glove wearing through performance feedback. Journal of Applied Behavior Analysis, 24, 705-712.
Figure 1, bottom panel, page 709

Serna, L. A., Schumaker, J. B., Sherman, J. A., & Sheldon, J. B. (1991). In-home generalization of social interactions in families of adolescents with behavior problems. Journal of Applied Behavior Analysis, 24, 733-746.
Figure 1, parent, sixth panel, page 741

Charlop, M. H., & Trasowech, J. E. (1991). Increasing autistic children's daily spontaneous speech. Journal of Applied Behavior Analysis, 24, 747-762.
Figure 1, bottom panel, imitated, page 754
Figure 2, bottom panel, incorrect, page 755
Figure 2, bottom panel, spontaneous, page 755

Sainato, D. M., Goldstein, H., & Strain, P. S. (1992). Effects of self-evaluation on preschool children's use of social interaction strategies with their classmates with autism. Journal of Applied Behavior Analysis, 25, 127-142.
Figure 1, middle panel, page 133

Ingham , P., & Greer, R. D. (1992). Changes in student and teacher responses in observed and generalized settings as a function of supervisor observations. Journal of Applied Behavior Analysis, 25, 153-164.
Figure 2, teacher C, page 160
Figure 3, classroom C, page 162

Ducharme, J. M., & Feldman, M. A. (1992). Comparison of staff training strategies to promote generalized teaching skills. Journal of Applied Behavior Analysis, 25, 165-180.
Figure 1, staff D, across-program generalization 2, page 171
Figure 4, staff P, across-program generalization 1, page 176

Ezell, H., & Goldstein, H. (1992). Teaching idiom comprehension to children with mental retardation. Journal of Applied Behavior Analysis, 25, 181-192.
Figure 1, bottom panel, idiomatic, page 186

Chandler, L. K., Fowler, S. A., & Lubeck, R. C. (1992). An analysis of the effects of multiple setting events on the social behavior of preschool children with special needs. Journal of Applied Behavior Analysis, 25, 249-264.
Figure 1, Sara, page 254

Goldstein, H., & Cisar, C. L. (1992). Promoting interaction during sociodramatic play: Teaching scripts to typical preschoolers and classmates with disabilities. Journal of Applied Behavior Analysis, 25, 265-280.
Figure 2, bottom panel, theme related, page 273

Kamps, D. M., Leonard, B. R., Vernon, S., Dugan, E. P., Delquadri, J. C., Gershon, B., Wade, L., & Folk, L. (1992). Teaching social skills to students with autism to increase peer interactions in an integrated first-grade classroom. Journal of Applied Behavior Analysis, 25, 281-288.
Figure 2, top panel, page 285

Odom, S. L., Chandler, L. K., Ostrosky, M., McConnell, S. R., & Reaney, S. (1992). Fading teacher prompts from peer-initiation interventions for young children with disabilities. Journal of Applied Behavior Analysis, 25, 307-318.
Figure 1, Quenton, peer initiations, page 312
Figure 1, Lee, peer initiations, page 312
Figure 1, Lee, teacher prompts, page 312
Figure 2, Quenton, page 313

Koegel, L. K., Koegel, R. L., Hurley, C., & Frea, W. D. (1992). Improving social skills and disruptive behavior in children with autism through self-management. Journal of Applied Behavior Analysis, 25, 341-354.
Figure 1, Adam, school/home, page 348
Figure 3, Ian, page 351

Kern-Dunlap, L., Dunlap, G., Clarke, S., Childs, K. E., White, R. L., & Stewart, M. P. (1992). Effects of a videotape feedback package on the peer interactions of children with serious behavioral and emotional challenges. Journal of Applied Behavior Analysis, 25, 355-364.
Figure 1, Dave, desirable, page 360

Morgan, R. L., & Salzberg, C. L. (1992). Effects of video-assisted training on employment-related social skills of adults with severe mental retardation. Journal of Applied Behavior Analysis, 25, 365-384.
Figure 1, water faucet on, video assessment, page 379
Figure 1, chair in the aisle, video assessment, page 379

Hundert, J., & Hopkins, B. (1992). Training supervisors in a collaborative team approach to promote peer interaction of children with disabilities in integrated preschools. Journal of Applied Behavior Analysis, 25, 385-400.
Figure 3, Vicky, training sessions, page 394
Figure 4, class 2, generalization sessions, page 396

Cuvo, A. J., Davis, P. K., O'Reilly, M. F., Mooney, B. M., & Crowley, R. (1992). Promoting stimulus control with textual prompts and performance feedback for persons with mild disabilities. Journal of Applied Behavior Analysis, 25, 477-490.
Figure 1, Mary, oven, page 481
Figure 2, Sara, refrigerator, page 482
Figure 3, Ann, refrigerator, page 484
Figure 6, Fran, refrigerator, page 487

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. Journal of Applied Behavior Analysis, 25, 491-498.
Figure 2, Ava, high-high, page 496

Harchik, A. E., Sherman, J. A., Sheldon, J. B., & Strouse, M. C. (1992). Ongoing consultation as a method of improving performance of staff members in a group home. Journal of Applied Behavior Analysis, 25, 599-610.
Figure 1, Karen, middle panel, page 606
Figure 2, Susie, middle panel, page 607
Figure 3, Pauline, middle panel, page 608
Figure 3, Pauline, bottom panel, page 608
Figure 4, Ernie, middle panel, page 609

Fleming, R., & Sulzer-Azaroff, B. (1992). Reciprocal peer management: Improving staff instruction in a vocational training program. Journal of Applied Behavior Analysis, 25, 611-620.

Figure 1, pair 3, Barbara, page 616
Figure 2, Jason's group, page 617

Allen, K. D., Loiben, T., Allen, S. J., & Stanley, R. T. (1992). Dentist-implemented contingent escape for management of disruptive child behavior. Journal of Applied Behavior Analysis, 25, 629-636.
Figure 1, Jenny, page 632

Hume, K. M., & Crossman, J. (1992). Musical reinforcement of practice behaviors among competitive swimmers. Journal of Applied Behavior Analysis, 25, 665-670.
Figure 1, Kevin, nonproductive, page 669

De Luca, R. V., & Holborn, S. W. (1992). Effects of a variable-ratio reinforcement schedule with changing criteria on exercise in obese and nonobese boys. Journal of Applied Behavior Analysis, 25, 671-680.
Figure 2, Steve, page 676

Wilson, P. G., Rusch, F. R., & Lee, S. (1992). Strategies to increase exercise-report correspondence by boys with moderate mental retardation: Collateral changes in intention-exercise correspondence. Journal of Applied Behavior Analysis, 25, 681-690.
Figure 1, Steve, promise-do, page 686
Figure 1, Jack, do-report, page 686

Repp, A. C., & Karsh, K. G. (1992). An analysis of a group teaching procedure for persons with developmental disabilities. Journal of Applied Behavior Analysis, 25, 701-712.
Figure 1, unprompted correct per minute, group 2, pate 707

Rasing, E. J., & Duker, P. C. (1992). Effects of a multifaceted training procedure on the acquisition and generalization of social behaviors in language-disabled deaf children. Journal of Applied Behavior Analysis, 25, 723-734.
Figure 1, middle panel, inappropriate instances, page 729
Figure 2, bottom panel, inappropriate instances, page 730

Durand, M. V., & Carr, E. G. (1992). An analysis of maintenance following functional communication training. Journal of Applied Behavior Analysis, 25, 777-794.
Figure 3, Jonathan, page 787
Figure 3, Dan, page 787
Figure 4, Ian, page 788

Charlop, M. H., Kurtz, P. F., & Milstein, J. P. (1992). Too much reinforcement, too little behavior: Assessing task interspersal procedures in conjunction with different reinforcement schedules with autistic children. Journal of Applied Behavior Analysis, 25, 795-808.
Figure 1, top panel, inappropriate behaviors, page 801
Figure 2, bottom panel, maintenance tasks, page 802
Figure 4, top panel, maintenance tasks, page 804

Cooper, L. J., Wacker, D. P., Thursby, D., Plagmann, L. A., Harding, J., Millard, T., & Derby, M. (1992). Analysis of the effects of task preferences, task demands, and adult attention on child behavior in outpatient and classroom settings. Journal of Applied Behavior Analysis, 25, 823-840.
Figure 3, bottom panel, HDLP, page 834

Fisher, W., Piazza, C., Cataldo, M., Harrell, R., Jefferson, G., & Conner, R. (1993). Functional communication training with and without extinction and punishment. Journal of Applied Behavior Analysis, 25, 23-36.
Figure 2, bottom panel, page 29

Wagaman, J. R., Miltenberger, R. G., & Arndorfer, R. E. (1993). Analysis of a simplified treatment for stuttering in children. Journal of Applied Behavior Analysis, 25, 53-62.
Figure 1, Nate, home observations, page 57
Figure 1, Pat, generalization probes, page 57

MacDuff, G. S., Krantz, P. J., & McClannahan, L. E. (1993). Teaching children with autism to use photographic activity schedules: Maintenance and generalization of complex response chains. Journal of Applied Behavior Analysis, 26, 89-98.
Figure 1, bottom panel, on-task, page 94

Krantz, P. J., & McClannahan, L. E. (1993). Teaching children with autism to initiate to peers: Effects of a script-fading procedure. Journal of Applied Behavior Analysis, 26, 121-132.
Figure 1, Walt, initiation, page 127

Derrickson, J. G., Neef, N. A., & Cataldo, M. F. (1993). Effects of signaling invasive procedures on a hospitalized infant's affective behaviors. Journal of Applied Behavior Analysis, 26, 133-134.
Case 2, PM, page 139

Carr, E. G., & Carlson, J. I. (1993). Reduction of severe behavior problems in the community using a multicomponent treatment approach. Journal of Applied Behavior Analysis, 26, 157-172.
Figure 1, top panel, page 167

Hendrickson, J. M., Gardner, N., Kaiser, A., & Riley, A. (1993). Evaluation of a social interaction coaching program in an integrated day-care setting. Journal of Applied Behavior Analysis, 26, 213-226.
Figure 2, bottom panel, page 221

Austin, J., Hatfield, D. B., Grindle, A. C., & Bailey, J. S. (1993). Increasing recycling in office environments: The effects of specific, informative cues. Journal of Applied Behavior Analysis, 26, 247-254.
Figure 1, top panel, page 250

Chapman, S., Fisher, W., Piazza, C. C., & Kurtz, P. F. (1993). Functional assessment and treatment of life-threatening drug ingestion in a dually diagnosed youth. Journal of Applied Behavior Analysis, 26, 255-256.
Bottom panel, page 256

Nau, P. A., Van Houten, R., Rolider, A., & Jonah, B. A. (1993). The failure of feedback on alcohol impairment to reduce impaired driving. Journal of Applied Behavior Analysis, 26, 361-368.
Figure 1, top panel, page 364
Figure 1, bottom panel, page 364

Koegel, R. L., & Frea, W. D. (1993). Treatment of social behavior in autism through the modification of pivotal social skills. Journal of Applied Behavior Analysis, 26, 369-378.
Figure 1, middle panel, page 374

Stock, L. Z., & Milan, M. A. (1993). Improving dietary practices of elderly individuals: The power of prompting, feedback, and social reinforcement. Journal of Applied Behavior Analysis, 26, 379-388.
Figure 1, bottom panel, page 383

Belfiore, P. J., Browder, D. M., & Mace, F. C. (1993). Effects of community and center-based settings on the alertness of persons with profound mental retardation. Journal of Applied Behavior Analysis, 26, 401-402.
Dave, percentage of intervals eye contact, low stimulation community, page 402
Nancy, percentage intervals head orienting, high-stimulation community, page 402

Werle, M. A., Murphy, T. B., & Budd, K. S. (1993). Treating chronic food refusal in young children: Home-based parent training. Journal of Applied Behavior Analysis, 26, 421-434.
Figure 1, positive attention, Bob, page 428
Figure 3, bottom panel, page 430

Stark, L. J., Knapp, L. G., Bowen, A. M., Powers, S. W., Jelalian, E., Evans, S., Passero, M. A., Mulvihill, M. M., & Hovell, M. (1993). Increasing calorie consumption in children with cystic fibrosis: Replication with 2-year follow-up. Journal of Applied Behavior Analysis, 26, 435-450.
Figure 2, Sam, pre-treatment, page 445

Repp, A. C., & Karsh, K. G. (1994). Hypothesis-based interventions for tantrum behaviors of persons with developmental disabilities in school settings. Journal of Applied Behavior Analysis, 27, 21-32.
Figure 2, middle panel, page 27

Kamps, D. M., Barbetta, P. M., Leonard, B. R., & Delquadri, J. (1994). Classwide peer tutoring: An integration strategy to improve reading skills and promote peer interactions among students with autism and general education peers. Journal of Applied Behavior Analysis, 27, 49-62.
Figure 1, middle panel, page 55
Figure 2, top panel, page 57
Figure 2, bottom panel, page 57

Miller, D. L., & Kelley, M. L. (1994). The use of goal setting and contingency contracting for improving children's homework performance. Journal of Applied Behavior Analysis, 27, 73-84.
Figure 2, bottom panel, page 80

Stoner, G., Carey, S. P., Ikeda, M. J., & Shinn, M. R. (1994). The utility of curriculum-based measurement for evaluating the effects of methylphenidate on academic performance. Journal of Applied Behavior Analysis, 27, 101-114.
Figure 2, bottom panel, page 109

Gillat, A., & Sulzer-Azaroff, B. (1994). Promoting principals' managerial involvement in instructional improvement. Journal of Applied Behavior Analysis, 27, 115-130.
Figure 4, 7th period, student 3, page 126
Figure 4, 7th period, student 6, page 126

Johnson, M. D., & Fawcett, S. B. (1994). Courteous service: Its assessment and modification in a human service organization. Journal of Applied Behavior Analysis, 27, 145-152.
Figure 1, top panel, page 150

Kennedy, C. H. (1994). Manipulating antecedent conditions to alter the stimulus control of problem behavior. Journal of Applied Behavior Analysis, 27, 161-170.
Figure 2, bottom panel, social affect, page 166

Goh, H., & Iwata, B. A. (1994). Behavioral persistence and variability during extinction of self-injury maintained by escape. Journal of Applied Behavior Analysis, 27, 173-174.
Bottom panel, page 174

Zarcone, J. R., Iwata, B. A., Smith, R. G., Mazaleski, J. L., & Lerman, D. C. (1994). Reemergence and extinction of self-injurious escape behavior during stimulus (instructional) fading. Journal of Applied Behavior Analysis, 27, 307-316.
Figure 2, middle panel, SIB, page 312

Rortvedt, A. K., & Miltenberger, R. G. (1994). Analysis of a high-probability instructional sequence and time-out in the treatment of child noncompliance. Journal of Applied Behavior Analysis, 27, 327-330.
Figure 1, bottom panel, page 329

Vollmer, T. R., Marcus, B. A., & LeBlanc, L. (1994). Treatment of self-injury and hand mouthing following inconclusive functional analyses. Journal of Applied Behavior Analysis, 27, 331-344.
Figure 2, top panel, page 337

Fisher, W. W., Piazza, C. C., Bowman, L. G., Kurtz, P. F., Sherer, M. R., & Lachman, S. R. (1994). A preliminary evaluation of empirically derived consequences for the treatment of pica. Journal of Applied Behavior Analysis, 27, 447-458.
Figure 1, Tom, page 454

Pierce, K. L., & Schreibman, L. (1994). Teaching daily living skills to children with autism in unsupervised settings through pictorial self-management. Journal of Applied Behavior Analysis, 27, 471-482.
Figure 3, bottom panel, inappropriate behavior, page 477

Farmer-Dougan, V. (1994). Increasing requests by adults with developmental disabilities using incidental teaching by peers. Journal of Applied Behavior Analysis, 27, 533-544.
Figure 1, bottom panel, page 538
Figure 2, bottom panel, page 539

Davis, C. A., Brady, M. P., Hamilton, R., McEvoy, M. A., & Williams, R. E. (1994). Effects of high-probability requests on the social interactions of young children with severe disabilities. Journal of Applied Behavior Analysis, 27, 619-638.
Figure 1, middle panel, low-P request, page 626

Ducharme, J. M., & Worling, D. E. (1994). Behavioral momentum and stimulus fading in the acquisition and maintenance of child compliance in the home. Journal of Applied Behavior Analysis, 27, 639-648.
Figure 1, bottom panel, page 644

Zarcone, J. R., Iwata, B. A., Mazaleski, J. L., & Smith, R. G. (1994). Momentum and extinction effects on self-injurious escape behavior and noncompliance. Journal of Applied Behavior Analysis, 27, 649-658.
Figure 2, middle panel, page 654
Figure 3, middle panel, page 656

Young, J. M., Krantz, P. J., McClannahan, L. E., & Poulson, C. L. (1994). Generalized imitation and response-class formation in children with autism. Journal of Applied Behavior Analysis, 27, 685-698.
Figure 1, Neal-Training Trials, pantomime response, matches, page 691
Figure 3, Heidi-Training Trials, pantomime response, non-matches, page 693
Figure 3, Heidi-Training Trials, pantomime response, matches, page 693
Figure 3, Heidi-Probe Trials, pantomime response, non-matches, page 693
Figure 4, Seth-Training Trials, vocal response, non-matches, page 695
Figure 4, Seth-Training Trials, pantomime response, non-matches, page 695

Taylor, B. A., & Harris, S. L. (1995). Teaching children with autism to seek information: Acquisition of novel information and generalization of responding. Journal of Applied Behavior Analysis, 28, 3-14.
Figure 2, bottom panel, receptive test, page 10

Kern, L., Wacker, D. P., Mace, F. C., Falk, G. D., Dunlap, G., & Kromrey, J. D. (1995). Improving the peer interactions of students with emotional and behavioral disorders through self-evaluation procedures: A component analysis and group application. Journal of Applied Behavior Analysis, 28, 47-60.
Figure 1, middle panel, inappropriate, page 51
Figure 2, bottom panel, appropriate, page 56

Smith, R. G., Iwata, B. A., & Reid, D. H. (1995). Effects of a voice output communication aid on interactions between support personnel and an individual with multiple disabilities. Journal of Applied Behavior Analysis, 28, 61-72.
Figure 1, bottom panel, page 66

Brigham, T. A., Meier, S. M., & Goodner, V. (1995). Increasing designated driving with a program of prompts and incentives. Journal of Applied Behavior Analysis, 28, 83-84.
Figure, page 84

Appendix B

List of Figures From Which the Functional Analysis (N=50)
Subsample of the 1990-1995 *JABA* Sample Was Drawn

For ease of retrieval, the following list of functional analysis baseline figures (N=50) and the articles that they were drawn from is organized according to the order in which they appear in the *Journal of Applied Behavior Analysis* beginning with Issue 1, Volume 23 (1990). Each article listed generally comprises more than one data set, and two or more data sets may have been drawn from the same article or figure, therefore each sample is identified with sufficient information to clearly distinguish it from other data sets within the figure or article.

Iwata, B. A., Pace, G. M., Kalsher, M. J., Cowdery, G. E., & Cataldo, M. F. (1990). Experimental analysis and extinction of self-injurious escape behavior. Journal of Applied Behavior Analysis, 23, 11-28.
Figure 1, Noreen, demand, page 16
Figure 1, Gina, demand, page 16
Figure 1, Gina, attention, page 16
Figure 1, Nellie, play, page 16
Figure 1, Donald, play, page 16
Figure 1, Donald, medical, page 16
Figure 1, Lana, play, page 16

Mason, S. A., & Iwata, B. A. (1990). Artifactual effects of sensory-integrative therapy on self-injurious behavior. Journal of Applied Behavior Analysis, 23, 361-370.
Figure 1, Kathy, demand, page 366
Figure 1, Kathy, attention, page 366

Steege, M. W., Wacker, D. P., Cigrand, K. C., Berg, W. K., Novak, C. G., Reamers, T. M., Sasso, G. M., & DeRaad, A. (1990). Use of negative reinforcement in the treatment of self-injurious behavior. Journal of Applied Behavior Analysis, 23, 459-468.
Figure 4, Ann, social attention, page 463

Cowdery, G. E., Iwata, B. A., & Pace, G. M. (1990). Effects and side effects of DRO as treatment for self-injurious behavior. Journal of Applied Behavior Analysis, 23, 497-506.
Figure 1, top panel, page 500

Mace, F. C., & Belfiore, P. (1990). Behavioral momentum in the treatment of escape-motivated stereotypy. Journal of Applied Behavior Analysis, 23, 507-514.
Figure 2, top panel, social disapproval, page 511

McKenzie, T. L., Sallis, J. F., Nader, P. R., Patterson, T. L., Elder, J. P., Berry, C. C., Rupp, J. W., Atkins, C. J., Buono, M. J., & Nelson, J. A. (1991). BEACHES: An observational system for assessing children's eating and physical activity behaviors and associated events. Journal of Applied Behavior Analysis, 24, 141-152.
Figure 1, high % bites graph, bites, page 147
Figure 1, low activity graph, activity prompts, page 147
Figure 1, low % bites graph, food prompts, page 147

Mace, F. C., & Lalli, J. S. (1991). Linking descriptive and experimental analyses in the treatment of bizarre speech. Journal of Applied Behavior Analysis, 24, 553-562.

Figure 1, middle panel, task disengagement, page 556

Smith, R. G., Iwata, B. A., Vollmer, T. R., & Pace, G. M. (1992). On the relationship between self-injurious behavior and self-restraint. Journal of Applied Behavior Analysis, 25, 433-446.

Figure 1, Lonnie, SIB alone, page 438
Figure 2, self restraint, alone, page 440
Figure 3, restraint unavailable, SIB play, page 441
Figure 4, SIB, play, page 443
Figure 4, alone, self-restraint, page 443
Figure 4, play, self-restraint, page 443

Vollmer, T. R., Iwata, B. A., Zarcone, J. R., Smith, R. G., & Mazaleski, J. L. (1993). The role of attention in the treatment of attention-maintained self-injurious behavior: Noncontingent reinforcement and differential reinforcement of other behavior. Journal of Applied Behavior Analysis, 26 9-22.

Figure 1, Brenda, alone, page 15
Figure 1, Bonnie, demand, page 15

Smith, R. G., Iwata, B. A., Vollmer, T. R., & Zarcone, J. R. (1993). Experimental analysis and treatment of multiply controlled self-injury. Journal of Applied Behavior Analysis, 26, 183-196.

Figure 1, Marc, attention, page 187
Figure 4, bottom panel, page 193

Chapman, S., Fisher, W., Piazza, C. C., & Kurtz, P. F. (1993). Functional assessment and treatment of life-threatening drug ingestion in a dually diagnosed youth. Journal of Applied Behavior Analysis, 26, 255-256.

Top panel, parent attention, page 256

Lerman, D. C., & Iwata, B. A. (1993). Descriptive and experimental analyses of variables maintaining self-injurious behavior. Journal of Applied Behavior Analysis, 26, 293-320.

Figure 1, Holly, no antecedent interaction (therapist present), page 302
Figure 1, Larry, antecedent instruction, page 302

Zarcone, J. R., Iwata, B. A., Vollmer, T. R., Jagtiani, S., Smith, R. G., & Mazaleski, J. L. (1993). Extinction of self-injurious escape behavior with and without instructional fading. Journal of Applied Behavior Analysis, 26, 353-360.

Figure 1, top panel, alone, page 357
Figure 1, top panel, play, page 357
Figure 2, bottom panel, SIB (extinction), page 358

Iwata, B. A., Pace, G. M., Cowdery, G. E., & Miltenberger, R. G. (1994). What makes extinction work: An analysis of procedural form and function. Journal of Applied Behavior Analysis, 27, 131-144.

Figure 2, bottom panel, page 137
Figure 3, demand, page 138

Figure 4, attention, page 139

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. Journal of Applied Behavior Analysis, 27, 197-210.
Figure 2, child 2, alone, page 205
Figure 2, child 5, social disapproval, page 205

Derby, K. M., Wacker, D. P., Peck, S., Sasso, G., DeRaad, A., Berg, W., Asmus, J., & Ulrich, S. (1994). Functional analysis of separate topographies of aberrant behavior. Journal of Applied Behavior Analysis, 27, 267-278.
Figure 1, top panel, diverted attention/ignore, page 273
Figure 1, middle panel, social attention, page 273
Figure 1, bottom panel, diverted attention/ignore, page 273

Pace, G. M., Ivancic, M. T., & Jefferson, G. (1994). Stimulus fading as treatment for obscenity in a brain-injured adult. Journal of Applied Behavior Analysis, 27, 301-306.
Figure 1, open triangles, page 303

Hagopian, L. P., Fisher, W. W., & Legacy, S. M. (1994). Schedule effects of noncontingent reinforcement on attention-maintained destructive behavior in identical quadruplets. Journal of Applied Behavior Analysis, 27, 317-326.
Figure 1, Wanda, demand, page 321
Figure 1, Laurie, demand, page 321
Figure 1, Lynn, demand, page 321

Lerman, D. C., Iwata, B. A., Smith, R. G., Zarcone, J. R., & Vollmer, T. R. (1994). Transfer of behavioral function as a contributing factor in treatment relapse. Journal of Applied Behavior Analysis, 27, 357-370.
Figure 1, top panel, alone, page 361
Figure 1, top panel, play, page 361

Zarcone, J. R., Iwata, B. A., Mazaleski, J. L., & Smith, R. G. (1994). Momentum and extinction effects on self-injurious escape behavior and noncompliance. Journal of Applied Behavior Analysis, 27, 649-658.
Figure 1, bottom panel, alone, page 653

Lalli, J. S., Casey, S. Goh, H., & Merlino, J. (1994). Treatment of escape-maintained aberrant behavior with escape extinction and predictable routines. Journal of Applied Behavior Analysis, 27, 705-714.
Figure 1, top panel, tangible, page 710
Figure 1, bottom panel escape, page 710

Vollmer, T. R., Marcus, B. A., & Ringdahl, J. E. (1995). Noncontingent escape as treatment for self-injurious behavior maintained by negative reinforcement. Journal of Applied Behavior Analysis, 28, 15-26.
Figure 1, top panel, escape, page 21

BIBLIOGRAPHY

*Alavosius, M. P., & Sulzer-Azaroff, B. (1990). Acquisition and maintenance of health-care routines as a function of feedback density. Journal of Applied Behavior Analysis, 23, 151-162.

*Allen, K. D., Loiben, T., Allen, S. J., & Stanley, R. T. (1992). Dentist-implemented contingent escape for management of disruptive child behavior. Journal of Applied Behavior Analysis, 25, 629-636.

*Austin, J., Hatfield, D. B., Grindle, A. C., & Bailey, J. S. (1993). Increasing recycling in office environments: The effects of specific, informative cues. Journal of Applied Behavior Analysis, 26, 247-254.

Baer, D. M. (1977). Perhaps it would be better not to know everything. Journal of Applied Behavior Analysis, 10, 167-172.

Baer, D. M., & Parsonson, B. S. (1978). Visual analysis of single-subject behavioral data. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change. (pp. 197-235). New York: Academic Press.

Baer, D. M., & Parsonson, B. S. (1992). The visual analysis of data , and current research into the stimuli controlling it. Single-case research design and analysis: New directions for psychology and education. Hillsdale, NJ: Lawrence Erlbaum Associates.

Barlow, D. H., & Hersen, M. (1984). Single case experimental designs: Strategies for studying behavior change. New York: Pergamon Press.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. Journal of the Royal Statistical Society, 8, 27-41.

*Belfiore, P. J., Browder, D. M., & Mace, F. C. (1993). Effects of community and center-based settings on the alertness of persons with profound mental retardation. Journal of Applied Behavior Analysis, 26, 401-402.

*Bell, K. E., Young, K. R., Salzberg, C. L., & West, R. P. (1991). High school driver education using peer tutors, direct instruction, and precision teaching. Journal of Applied Behavior Analysis, 24, 45-52.

Box, G. E. P., & Jenkins G. M. (1970). Time-series analysis: Forecasting and control. San Francisco: Holden-Day.

---

* Asterisk marks references for articles contained the 1990-1995 *JABA* sample and subsamples used in the meta-analysis

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis, forecasting and control (Revised Edition). San Francisco: Holden-Day.

Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated-moving average time-series models. Journal of the American Statistical Association, 65, 1509-1526.

Box, G. E. P., & Tiao, G. C. (1965). A change in level of a nonstationary time series. Biometrika, 52, 181-192.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 70, 70-79.

*Brigham, T. A., Meier, S. M., & Goodner, V. (1995). Increasing designated driving with a program of prompts and incentives. Journal of Applied Behavior Analysis, 28, 83-84

*Burgio, L. D., Engel, B. T., Hawkins, A., McCormick, K., Scheve, A., & Jones, L. T. (1990). A staff management system for maintaining improvements in continence with elderly nursing home residents. Journal of Applied Behavior Analysis, 23, 111-118.

Busk, P. L., & Marascuilo, L. A. (1988). Autocorrelation in single-subject research: A counter argument to the myth of no autocorrelation. Behavioral Assessment, 10, 229-242.

*Carr, E. G., & Carlson, J. I. (1993). Reduction of severe behavior problems in the community using a multicomponent treatment approach. Journal of Applied Behavior Analysis, 26, 157-172.

*Chandler, L. K., Fowler, S. A., & Lubeck, R. C. (1992). An analysis of the effects of multiple setting events on the social behavior of preschool children with special needs. Journal of Applied Behavior Analysis, 25, 249-264.

*Chapman, S., Fisher, W., Piazza, C. C., & Kurtz, P. F . (1993). Functional assessment and treatment of life-threatening drug ingestion in a dually diagnosed youth. Journal of Applied Behavior Analysis, 26, 255-256.

*Charlop, M. H., Kurtz, P. F., & Casey, F. G. (1990). Using aberrant behaviors as reinforcers for autistic children. Journal of Applied Behavior Analysis, 23, 163-182.

*Charlop, M. H., Kurtz, P. F., & Milstein, J. P. (1992). Too much reinforcement, too little behavior: Assessing task interspersal procedures in conjunction with different reinforcement schedules with autistic children. Journal of Applied Behavior Analysis, 25, 795-808.

*Charlop, M. H., & Trasowech, J. E. (1991). Increasing autistic children's daily spontaneous speech. Journal of Applied Behavior Analysis, 24, 747-762.

*Cooper, L. J., Wacker, D. P., Sasso, G. M., Reimers, T. M., & Donn, L. K. (1990). Using parents as therapists to evaluate appropriate behavior of their children: Application to a tertiary diagnostic clinic. Journal of Applied Behavior Analysis, 23, 285-296.

*Cooper, L. J., Wacker, D. P., Thursby, D., Plagmann, L. A., Harding, J., Millard, T., & Derby, M. (1992). Analysis of the effects of task preferences, task demands, and adult attention on child behavior in outpatient and classroom settings. Journal of Applied Behavior Analysis, 25, 823-840.

*Cope, J. G., & Allred, L. J. (1991). Community intervention to deter illegal parking in spaces reserved for the physically disabled. Journal of Applied Behavior Analysis, 24, 687-694.

*Cowdery, G. E., Iwata, B. A., & Pace, G. M. (1990). Effects and side effects of DRO as treatment for self-injurious behavior. Journal of Applied Behavior Analysis, 23, 497-506.

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

*Cuvo, A. J., Davis, P. K., O'Reilly, M. F., Mooney, B. M., & Crowley, R. (1992). Promoting stimulus control with textual prompts and performance feedback for persons with mild disabilities. Journal of Applied Behavior Analysis, 25, 477-490.

*Dattilo, J., & Camarata, S. (1991). Facilitating conversation through self-initiated augmentative communication treatment. Journal of Applied Behavior Analysis, 24, 369-378.

*Davis, C. A., Brady, M. P., Hamilton, R., McEvoy, M. A., & Williams, R. E. (1994). Effects of high-probability requests on the social interactions of young children with severe disabilities. Journal of Applied Behavior Analysis, 27, 619-638.

*De Luca, R. V., & Holborn, S. W. (1992). Effects of a variable-ratio reinforcement schedule with changing criteria on exercise in obese and nonobese boys. Journal of Applied Behavior Analysis, 25, 671-680.

DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. Journal of Applied Behavior Analysis, 12, 573-579.

*Derby, K. M., Wacker, D. P., Peck, S., Sasso, G., DeRaad, A., Berg, W., Asmus, J., & Ulrich, S. (1994). Functional analysis of separate topographies of aberrant behavior. Journal of Applied Behavior Analysis, 27, 267-278.

*Derrickson, J. G., Neef, N. A., & Cataldo, M. F. (1993). Effects of signaling invasive procedures on a hospitalized infant's affective behaviors. Journal of Applied Behavior Analysis, 26, 133-134.

*Derrickson, J. G., Neef, N. A., & Parrish, J. M. (1991). Teaching self-administration of suctioning to children with tracheostomies. Journal of Applied Behavior Analysis, 24, 563-570.

*DeVries, J. E., Burnette, M. M., & Redmon, W. K. (1991). AIDS prevention: Improving nurses' compliance with glove wearing through performance feedback. Journal of Applied Behavior Analysis, 24, 705-712.

*Ducharme, J. M., & Feldman, M. A. (1992). Comparison of staff training strategies to promote generalized teaching skills. Journal of Applied Behavior Analysis, 25, 165-180.

*Ducharme, J. M., & Worling, D. E. (1994). Behavioral momentum and stimulus fading in the acquisition and maintenance of child compliance in the home. Journal of Applied Behavior Analysis, 27, 639-648.

*Duker, P. C., & van Lent, C. (1991). Inducing variability in communicative gestures used by severely retarded individuals. Journal of Applied Behavior Analysis, 24, 379-386.

*Durand, M. V., & Carr, E. G. (1991). Functional communication training to reduce challenging behavior: Maintenance and application in new settings. Journal of Applied Behavior Analysis, 24, 251-264.

*Durand, M. V., & Carr, E. G. (1992). An analysis of maintenance following functional communication training. Journal of Applied Behavior Analysis, 25, 777-794.

*Ezell, H., & Goldstein, H. (1992). Teaching idiom comprehension to children with mental retardation. Journal of Applied Behavior Analysis, 25, 181-192.

*Farmer-Dougan, V. (1994). Increasing requests by adults with developmental disabilities using incidental teaching by peers. Journal of Applied Behavior Analysis, 27, 533-544.

*Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. Journal of Applied Behavior Analysis, 25, 491-498.

*Fisher, W. W., Piazza, C. C., Bowman, L. G., Kurtz, P. F., Sherer, M. R., & Lachman, S. R. (1994). A preliminary evaluation of empirically derived consequences for the treatment of pica. Journal of Applied Behavior Analysis, 27, 447-458.

*Fisher, W., Piazza, C., Cataldo, M., Harrell, R., Jefferson, G., & Conner, R. (1993). Functional communication training with and without extinction and punishment. Journal of Applied Behavior Analysis, 26, 23-36.

*Fleming, R., & Sulzer-Azaroff, B. (1992). Reciprocal peer management: Improving staff instruction in a vocational training program. Journal of Applied Behavior Analysis, 25, 611-620.

*France, K. G., & Hudson, S. M. (1990). Behavior management of infant sleep disturbance. Journal of Applied Behavior Analysis, 23, 91-98.

Gentile, J. R., Roden, A. H., & Klein, R. D. (1962). An analysis-of-variance-model for the intrasubject replication design. Journal of Applied Behavior Analysis, 5, 193-198.

*Gillat, A., & Sulzer-Azaroff, B. (1994). Promoting principals' managerial involvement in instructional improvement. Journal of Applied Behavior Analysis, 27, 115-130.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). Design and analysis of time-series experiments. Boulder, Colorado: Colorado University Press.

*Goh, H., & Iwata, B. A. (1994). Behavioral persistence and variability during extinction of self-injury maintained by escape. Journal of Applied Behavior Analysis, 27, 173-174.

*Goldstein, H., & Cisar, C. L. (1992). Promoting interaction during sociodramatic play: Teaching scripts to typical preschoolers and classmates with disabilities. Journal of Applied Behavior Analysis, 25, 265-280.

Gottman, J. M., & Glass, G. V. (1978). Analysis of interrupted time-series experiments. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change. (pp. 197-235). New York: Academic Press.

*Green, C. W., Reid, D. H., Canipe, V. S., & Gardner, S. M. (1991). A comprehensive evaluation of reinforcer identification processes for persons with profound multiple handicaps. Journal of Applied Behavior Analysis, 24, 537-552.

*Green, C. W., Reid, D. H., Perkins, L., & Gardner, S. M. (1991). Increasing habilitative services for persons with profound handicaps: An application of structural analysis to staff management. Journal of Applied Behavior Analysis, 24, 459-472.

Greenwood, K. M., & Matyas, T. A. (1990). Problems with the application of interrupted time series for brief single-subject data. Behavioral Assessment, 12, 355-370.

*Hagopian, L. P., Fisher, W. W., & Legacy, S. M. (1994). Schedule effects of noncontingent reinforcement on attention-maintained destructive behavior in identical quadruplets. Journal of Applied Behavior Analysis, 27, 317-326.

*Harchik, A. E., Sherman, J. A., Sheldon, J. B., & Strouse, M. C. (1992). Ongoing consultation as a method of improving performance of staff members in a group home. Journal of Applied Behavior Analysis, 25, 599-610.

*Haring, T. G., & Kennedy, C. H. (1990). Contextual control of problem behavior in students with severe disabilities. Journal of Applied Behavior Analysis, 23, 235-244.

Hartmann, D. P., Gottman, J. M., Jones, R. R., Gardner, W., Kazdin, A. E., & Vaught, R. S. (1980). Interrupted time-series analysis and its application to behavioral data. Journal of Applied Behavior Analysis, 13, 543-559.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

*Hendrickson, J. M., Gardner, N., Kaiser, A., & Riley, A. (1993). Evaluation of a social interaction coaching program in an integrated day-care setting. Journal of Applied Behavior Analysis, 26, 213-226.

*Homer, R. H., Day, H. M., Sprague, J. R., O'Brien, M., & Heathfield, L. T. (1991). Interspersed requests: A nonaversive procedure for reducing aggression and self-injury during instruction. Journal of Applied Behavior Analysis, 24, 265-278.

Horne, B. P., Yang, S. P., & Ware, W. (1982). Interrupted tome-series analysis applied to small-N data series. Psychological Bulletin, 91, 178-189.

Huitema, B. E. (1985). Autocorrelation in applied behavior analysis: A myth. Behavioral Assessment, 7, 109-120.

Huitema, B. E. (1986a). Autocorrelation in behavioral research: Wherefore art thou? In A. Poling & R. W. Fuqua (Eds.) Research methods in applied behavior analysis: Issues and advances. New York: Plenum Press.

Huitema, B. E. (1986b). Statistical analysis and single-subject designs: Some misunderstandings. In A. Poling & R. W. Fuqua (Eds.) Research methods in applied behavior analysis: Issues and advances. New York: Plenum Press.

Huitema, B. E. (1988). Autocorrelation: 10 years of confusion. Behavioral Assessment, 10, 253-294.

Huitema, B. E. (1990, May). A secondary analysis of serial dependency. Invited address delivered at the meeting of the Association for Behavior Analysis, Nashville.

Huitema, B. E. (1994). Time-series analysis. In A. Corsini (Ed.) Wiley encyclopedia of psychology: 2nd edition. New York: John Wiley & Sons.

Huitema, B. E., & McKean, J. (1991). Autocorrelation estimation and inference with small samples. Psychological Bulletin, 110, 291-304.

Huitema, B. E., & McKean, J. W. (1994a). Reduced bias autocorrelation estimation: Three jackknife methods. Educational and Psychological Measurement, 54, 654-665.

Huitema, B. E., & McKean, J. W. (1994b). Two reduced-bias autocorrelation estimators: r_{F1} and r_{F2}. Perceptual and Motor Skills, 78, 323-330.

Huitema, B. E., & McKean, J. W. (1994c). Tests of H_0: $\rho_1 = 0$ for autocorrelation estimators r_{F1} and r_{F2}. Perceptual and Motor Skills, 78, 331-336.

Huitema, B. E., & McKean, J. (1994d) Some issues in small-sample time series intervention analysis. Presented at the meeting of the Association for Behavior Analysis, Atlanta.

Huitema, B. E., & McKean, J. W. (1995). Tests for the Jackknife Autocorrelation Estimator . Educational and Psychological Measurement.

Huitema, B. E., & McKean, J. W. (in press). Tests for the Jackknife Autocorrelation Estimator rQ2. Educational and Psychological Measurement, 54, 654-665

Huitema, B. E., McKean, J. W. & McKnight, S. (1994, June) New methods of intervention analysis: Simple and Complex. Presented at the meeting of the Association for Behavior Analysis, Atlanta.

Huitema, B. E., McKean, J. W., & McKnight, S. (1994, August). Small-sample time-series intervention analysis: Problems and solutions. Paper presented at the meeting of the American Psychological Association Meetings, Los Angeles.

Huitema, B. E., McKean, J. W. & Zhao, J. (in press). The runs test for autocorrelated errors: Unacceptable properties. Journal of Educational and Behavioral Statistics.

Huitema, B. E., McKean, J. W. & Zhao, J. (in preparation). A test for homogeneity of autocorrelation coefficients.

Huitema, B. E., McKean, J. W. & Zhao, J. (submitted). Durbin-Watson test of autocorrelated errors in time-series regression: Some problems.

*Hume, K. M., & Crossman, J. (1992). Musical reinforcement of practice behaviors among competitive swimmers. Journal of Applied Behavior Analysis, 25, 665-670.

*Hundert, J., & Hopkins, B. (1992). Training supervisors in a collaborative team approach to promote peer interaction of children with disabilities in integrated preschools. Journal of Applied Behavior Analysis, 25, 385-400.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-analysis: Cumulating findings across research. Beverly Hills, CA: Sage.

Huyser, K., & van der Laan, J. (1992). DataThief. Author.

*Ingenmey, R., & Van Houten, R. (1991). Using time delay to promote spontaneous speech in an autistic child. Journal of Applied Behavior Analysis, 24, 591-596.

*Ingham , P., & Greer, R. D. (1992). Changes in student and teacher responses in observed and generalized settings as a function of supervisor observations. Journal of Applied Behavior Analysis, 25, 153-164.

*Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. Journal of Applied Behavior Analysis, 27, 197-210.

*Iwata, B. A., Pace, G. M., Cowdery, G. E., & Miltenberger, R. G. (1994). What makes extinction work: An analysis of procedural form and function. Journal of Applied Behavior Analysis, 27, 131-144.

Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., Rodgers, T. A., Lerman, D. C., Shore, B. A., Mazaleski, J. L., Goh, H., Cowdery, G. E., Kalsher, M. J., McCosh, K. C., & Willis, K. D. (1994). The functions of self-injurious behavior: An experimental-epidemiological analysis. Journal of Applied Behavior Analysis, 2, 215-240.

*Iwata, B. A., Pace, G. M., Kalsher, M. J., Cowdery, G. E., & Cataldo, M. F. (1990). Experimental analysis and extinction of self-injurious escape behavior. Journal of Applied Behavior Analysis, 23, 11-28.

Johnson, B. T., Mullen, B., & Salas, E. (1995). Comparison of three major meta-analytic approaches. Journal of Applied Psychology, 80, 94-106.

*Johnson, M. D., & Fawcett, S. B. (1994). Courteous service: Its assessment and modification in a human service organization. Journal of Applied Behavior Analysis, 27, 145-152.

Johnston, J. M., & Pennypacker, H. S. (1991). Strategies and tactics of human behavioral research. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

*Johnston, J. M., Greene, K. S., Rawal, A., Vazin, T., & Winston, M. (1991). Effects of caloric level on ruminating. Journal of Applied Behavior Analysis, 24, 597-603.

Jones, R. R., Vaught, R. S., & Reid, J. B. (1975). Time series analysis as a substitute for single subject analysis of variance designs. In G. R. Patterson, I. M. Marks, J. D. Matarazzo, R. A. Myers, G. E. Schwartz, & H. H. Strupp (Eds.), Behavior change 1974. Hawthorne, NY: Aldine.

Jones, R. R., Vaught, R. S., & Weinrott, M. (1977). Time-series analysis in operant research. Journal of Applied Behavior Analysis, 10, 151-166.

Jones, R. R., Weinrott, M., & Vaught, R. S. (1978). Effects of serial dependency on the agreement between visual and statistical inferences. Journal of Applied Behavior Analysis, 11, 277-284.

*Kamps, D. M., Barbetta, P. M., Leonard, B. R., & Delquadri, J. (1994). Classwide peer tutoring: An integration strategy to improve reading skills and promote peer interactions among students with autism and general education peers. Journal of Applied Behavior Analysis, 27, 49-62.

*Kamps, D. M., Leonard, B. R., Vernon, S., Dugan, E. P., Delquadri, J. C., Gershon, B., Wade, L., & Folk, L. (1992). Teaching social skills to students with autism to increase peer interactions in an integrated first-grade classroom. Journal of Applied Behavior Analysis, 25, 281-288.

Kazdin, A. E. (1982). Single-case research designs: Methods for clinical and applied settings. New York: Oxford University Press.

*Kennedy, C. H. (1994). Manipulating antecedent conditions to alter the stimulus control of problem behavior. Journal of Applied Behavior Analysis, 27, 161-170.

*Kern, L., Wacker, D. P., Mace, F. C., Falk, G. D., Dunlap, G., & Kromrey, J. D. (1995). Improving the peer interactions of students with emotional and behavioral disorders through self-evaluation procedures: A component analysis and group application. Journal of Applied Behavior Analysis, 28, 47-60.

*Kern-Dunlap, L., Dunlap, G., Clarke, S., Childs, K. E., White, R. L., & Stewart, M. P. (1992). Effects of a videotape feedback package on the peer interactions of children with serious behavioral and emotional challenges. Journal of Applied Behavior Analysis, 25, 355-364.

*Kirby, K. C., Fowler, S. A., & Baer, D. M. (1991). Reactivity in self-recording: Obtrusiveness of recording procedure and peer comments. Journal of Applied Behavior Analysis, 24, 487-498.

*Koegel, L. K., Koegel, R. L., Hurley, C., & Frea, W. D. (1992). Improving social skills and disruptive behavior in children with autism through self-management. Journal of Applied Behavior Analysis, 25, 341-354.

*Koegel, R. L., & Frea, W. D. (1993). Treatment of social behavior in autism through the modification of pivotal social skills. Journal of Applied Behavior Analysis, 26, 369-378.

*Koegel, R. L., & Koegel, L. K. (1990). Extended reductions in stereotypic behavior of students with autism through a self-management treatment package. Journal of Applied Behavior Analysis, 23, 119-128.

*Kohler, F. W., & Greenwood, C. R. (1990). Effects of collateral peer supportive behaviors within the classwide peer tutoring program. Journal of Applied Behavior Analysis, 23, 307-322.

*Krantz, P. J., & McClannahan, L. E. (1993). Teaching children with autism to initiate to peers: Effects of a script-fading procedure. Journal of Applied Behavior Analysis, 26, 121-132.

*Lalli, J. S., Casey, S. Goh, H., & Merlino, J. (1994). Treatment of escape-maintained aberrant behavior with escape extinction and predictable routines. Journal of Applied Behavior Analysis, 27, 705-714.

*Lerman, D. C., & Iwata, B. A. (1993). Descriptive and experimental analyses of variables maintaining self-injurious behavior. Journal of Applied Behavior Analysis, 26, 293-320.

*Lerman, D. C., Iwata, B. A., Smith, R. G., Zarcone, J. R., & Vollmer, T. R. (1994). Transfer of behavioral function as a contributing factor in treatment relapse. Journal of Applied Behavior Analysis, 27, 357-370.

Liu, M. L. (1993). Modeling and forecasting time series using an expert system approach. Working Paper No. 127. Chicago: Scientific Computing Associates.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time-series models. Biometrika, 65, 297-303.

*Lombard, D., Neubauer, T. E., Canfield, D., & Winett, R. A. (1991). Behavioral community intervention to reduce the risk of skin cancer. Journal of Applied Behavior Analysis, 24, 677-686.

*Love, S. R., Matson, J. L., & West, D. (1990). Mothers as effective therapists for autistic children's phobias. Journal of Applied Behavior Analysis, 23, 379-386.

*Ludwig, T. D., & Geller, E. S. (1991). Improving the driving practices of pizza deliverers: Response generalization and moderating effects of driving history. Journal of Applied Behavior Analysis, 24, 31-44.

*MacDuff, G. S., Krantz, P. J., & McClannahan, L. E. (1993). Teaching children with autism to use photographic activity schedules: Maintenance and generalization of complex response chains. Journal of Applied Behavior Analysis, 26, 89-98.

*Mace, F. C., & Belfiore, P. (1990). Behavioral momentum in the treatment of escape-motivated stereotypy. Journal of Applied Behavior Analysis, 23, 507-514.

*Mace, F. C., & Lalli, J. S. (1991). Linking descriptive and experimental analyses in the treatment of bizarre speech. Journal of Applied Behavior Analysis, 24, 553-562.

*Mason, S. A., & Iwata, B. A. (1990). Artifactual effects of sensory-integrative therapy on self-injurious behavior. Journal of Applied Behavior Analysis, 23, 361-370.

*Matson, J. L., Sevin, J. A., Fridley, D., & Love, S. R. (1990). Increasing spontaneous language in three autistic children. Journal of Applied Behavior Analysis, 23, 227-234.

Matyas, T. A., & Greenwood, K. M. (1990). The effect of serial dependence on visual judgement of single-case charts: An addendum. The Occupational Therapy Journal of Research, 10, 208-219.

Matyas, T. A., & Greenwood, K. M. (1991). Problems in the estimation of autocorrelation in brief time series and some implications for behavioral data. Behavioral Assessment, 13, 137-157.

*McClannahan, L. E., McGee, G. G., MacDuff, G. S., & Krantz, P. J. (1990). Assessing and improving child care: A personal appearance index for children with autism. Journal of Applied Behavior Analysis, 23, 469-482.

McKean, J. W., & Huitema, B. E. (1993). Small sample properties of the Spearman autocorrelation estimator. Perceptual and Motor Skills, 76, 384-386.

*McKenzie, T. L., Sallis, J. F., Nader, P. R., Patterson, T. L., Elder, J. P., Berry, C. C., Rupp, J. W., Atkins, C. J., Buono, M. J., & Nelson, J. A. (1991). BEACHES: An observational system for assessing children's eating and physical activity behaviors and associated events. Journal of Applied Behavior Analysis, 24, 141-152.

McKnight, S. D. (1994). A bootstrap method to analyze an intervention model with autoregressive error terms. Unpublished doctoral dissertation, Western Michigan University.

McKnight, S. D., McKean, J. W., & Huitema, B. E. (in preparation). A bootstrap based time-series intervention model for small samples.

Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse Journal of Applied Behavior Analysis, 7, 647-653.

*Miller, D. L., & Kelley, M. L. (1994). The use of goal setting and contingency contracting for improving children's homework performance. Journal of Applied Behavior Analysis, 27, 73-84.

*Morgan, R. L., & Salzberg, C. L. (1992). Effects of video-assisted training on employment-related social skills of adults with severe mental retardation. Journal of Applied Behavior Analysis, 25, 365-384.

*Nau, P. A., Van Houten, R., Rolider, A., & Jonah, B. A. (1993). The failure of feedback on alcohol impairment to reduce impaired driving. Journal of Applied Behavior Analysis, 26, 361-368.

*Odom, S. L., Chandler, L. K., Ostrosky, M., McConnell, S. R., & Reaney, S. (1992). Fading teacher prompts from peer-initiation interventions for young children with disabilities. Journal of Applied Behavior Analysis, 25, 307-318.

*Osborne, K., Rudrud, E., & Zezoney, F. (1990). Improved curveball hitting through the enhancement of visual cues. Journal of Applied Behavior Analysis, 23, 371-378.

Ottenbacher, K. J. (1986). Reliability and accuracy of visually analyzing graphed data from single-subject designs. American Journal of Occupational Therapy, 40, 464-469.

*Pace, G. M., Ivancic, M. T., & Jefferson, G. (1994). Stimulus fading as treatment for obscenity in a brain-injured adult. Journal of Applied Behavior Analysis, 27, 301-306.

*Parsons, M. B., & Reid, D. H. (1990). Assessing food preferences among persons with profound mental retardation: Providing opportunities to make choices. Journal of Applied Behavior Analysis, 23, 183-196.

Parsonson, B. S., & Baer, D. M. (1986). The graphic analysis of data. In A. Poling & R. W. Fuqua (Eds.) Research methods in applied behavior analysis: Issues and advances. New York: Plenum Press.

*Piazza, C. C., & Fisher, W. (1991). A faded bedtime with response cost protocol for treatment of multiple sleep problems in children. Journal of Applied Behavior Analysis, 24, 129-140.

*Pierce, K. L., & Schreibman, L. (1994). Teaching daily living skills to children with autism in unsupervised settings through pictorial self-management. Journal of Applied Behavior Analysis, 27, 471-482.

*Ragnarsson R. S., & Bjorgvinsson, T. (1991). Effects of public posting on driving speed in Icelandic traffic. Journal of Applied Behavior Analysis, 24, 53-58.

*Rasing, E. J., & Duker, P. C. (1992). Effects of a multifaceted training procedure on the acquisition and generalization of social behaviors in language-disabled deaf children. Journal of Applied Behavior Analysis, 25, 723-734.

Raudenbush, S. W. (1991). [Review of the book Methods of meta-analysis: Correcting error and bias in research findings]. Journal of the American Statistical Association, March, 242-244.

*Repp, A. C., & Karsh, K. G. (1992). An analysis of a group teaching procedure for persons with developmental disabilities. Journal of Applied Behavior Analysis, 25, 701-712.

*Repp, A. C., & Karsh, K. G. (1994). Hypothesis-based interventions for tantrum behaviors of persons with developmental disabilities in school settings. Journal of Applied Behavior Analysis, 27, 21-32.

*Rortvedt, A. K., & Miltenberger, R. G. (1994). Analysis of a high-probability instructional sequence and time-out in the treatment of child noncompliance. Journal of Applied Behavior Analysis, 27, 327-330.

Rosenthal, R. (1991). Meta-analytic procedures for social research (Rev. ed.). Beverly Hills, CA: Sage.

*Sainato, D. M., Goldstein, H., & Strain, P. S. (1992). Effects of self-evaluation on preschool children's use of social interaction strategies with their classmates with autism. Journal of Applied Behavior Analysis, 25, 127-142.

*Serna, L. A., Schumaker, J. B., Sherman, J. A., & Sheldon, J. B. (1991). In-home generalization of social interactions in families of adolescents with behavior problems. Journal of Applied Behavior Analysis, 24, 733-746.

Sharpley, C. F. (1987). Time-series analysis of behavioral data: An update. Behavior Change, 4, 40-45.

Sharpley, C. F. (1988). Single-subject research. In Educational research, methodology, and measurement: An international handbook. John P. Reeves (Ed.) New York: Pergamon Press.

Sharpley, C. F., & Alavosius, M. P. (1988). Autocorrelation in behavioral data: An alternative perspective. Behavioral Assessment, 10, 243-251.

Sidman, M. (1960). Tactics of scientific research: Evaluating experimental data in psychology. New York: Basic Books, Inc.

*Smith, R. G., Iwata, B. A., & Reid, D. H. (1995). Effects of a voice output communication aid on interactions between support personnel and an individual with multiple disabilities. Journal of Applied Behavior Analysis, 28, 61-72.

*Smith, R. G., Iwata, B. A., Vollmer, T. R., & Pace, G. M. (1992). On the relationship between self-injurious behavior and self-restraint. Journal of Applied Behavior Analysis, 25, 433-446.

*Smith, R. G., Iwata, B. A., Vollmer, T. R., & Zarcone, J. R. (1993). Experimental analysis and treatment of multiply controlled self-injury. Journal of Applied Behavior Analysis, 26, 183-196.

*Sosland, J. M., & Christophersen, E. R. (1991). Does SleepTight work? A behavioral analysis of the effectiveness of SleepTight for the management of infant colic. Journal of Applied Behavior Analysis, 24, 161-166.

*Stark, L. J., Knapp, L. G., Bowen, A. M., Powers, S. W., Jelalian, E., Evans, S., Passero, M. A., Mulvihill, M. M., & Hovell, M. (1993). Increasing calorie consumption in children with cystic fibrosis: Replication with 2-year follow-up. Journal of Applied Behavior Analysis, 26, 435-450.

*Steege, M. W., Wacker, D. P., Cigrand, K. C., Berg, W. K., Novak, C. G., Reamers, T. M., Sasso, G. M., & DeRaad, A. (1990). Use of negative reinforcement in the treatment of self-injurious behavior. Journal of Applied Behavior Analysis, 23, 459-468.

*Stock, L. Z., & Milan, M. A. (1993). Improving dietary practices of elderly individuals: The power of prompting, feedback, and social reinforcement. Journal of Applied Behavior Analysis, 26, 379-388.

*Stoner, G., Carey, S. P., Ikeda, M. J., & Shinn, M. R. (1994). The utility of curriculum-based measurement for evaluating the effects of methylphenidate on academic performance. Journal of Applied Behavior Analysis, 27, 101-114.

Suen, H. K. (1987). On the epistemology of autocorrelation in applied behavior analysis. Behavioral Assessment, 9, 113-124.

Suen, H. K., & Ary, D. (1987). Autocorrelation in applied behavior analysis: Myth or reality? Behavioral Assessment, 9, 125-130.

Suen, H. K., Lee, P. S. C., & Owen, S, V. (1990). Effects of autocorrelation on single-subject single-facet crossed-design generalizability assessment. Behavioral Assessment, 12, 305-315.

*Taylor, B. A., & Harris, S. L. (1995). Teaching children with autism to seek information: Acquisition of novel information and generalization of responding. Journal of Applied Behavior Analysis, 28, 3-14.

Velicer, W. F., & McDonald, R. P. (1984). Time series analysis without model identification. Multivariate behavioral research, 19, 33-47.

*Vollmer, T. R., Iwata, B. A., Zarcone, J. R., Smith, R. G., & Mazaleski, J. L. (1993). The role of attention in the treatment of attention-maintained self-injurious behavior: Noncontingent reinforcement and differential reinforcement of other behavior. Journal of Applied Behavior Analysis, 26, 9-22.

*Vollmer, T. R., Marcus, B. A., & LeBlanc, L. (1994). Treatment of self-injury and hand mouthing following inconclusive functional analyses. Journal of Applied Behavior Analysis, 27, 331-344.

*Vollmer, T. R., Marcus, B. A., & Ringdahl, J. E. (1995). Noncontingent escape as treatment for self-injurious behavior maintained by negative reinforcement. Journal of Applied Behavior Analysis, 28, 15-26.

*Wagaman, J. R., Miltenberger, R. G., & Arndorfer, R. E. (1993). Analysis of a simplified treatment for stuttering in children. Journal of Applied Behavior Analysis, 26, 53-62.

*Ward, W. D., & Stare, S. W. (1990). The role of subject verbalization in generalized correspondence. Journal of Applied Behavior Analysis, 23, 129-137.

*Werle, M. A., Murphy, T. B., & Budd, K. S. (1993). Treating chronic food refusal in young children: Home-based parent training. Journal of Applied Behavior Analysis, 26, 421-434.

*White, A. G., & Bailey, J. S. (1990). Reducing disruptive behaviors of elementary physical education students with Sit and Watch. Journal of Applied Behavior Analysis, 23, 353-360.

*Wilson, P. G., Rusch, F. R., & Lee, S. (1992). Strategies to increase exercise-report correspondence by boys with moderate mental retardation: Collateral changes in intention-exercise correspondence. Journal of Applied Behavior Analysis, 25, 681-690.

*Young, J. M., Krantz, P. J., McClannahan, L. E., & Poulson, C. L. (1994). Generalized imitation and response-class formation in children with autism. Journal of Applied Behavior Analysis, 27, 685-698.

*Zarcone, J. R., Iwata, B. A., Mazaleski, J. L., & Smith, R. G. (1994). Momentum and extinction effects on self-injurious escape behavior and noncompliance. Journal of Applied Behavior Analysis, 27, 649-658.

*Zarcone, J. R., Iwata, B. A., Smith, R. G., Mazaleski, J. L., & Lerman, D. C. (1994). Reemergence and extinction of self-injurious escape behavior during stimulus (instructional) fading. Journal of Applied Behavior Analysis, 27, 307-316.

*Zarcone, J. R., Iwata, B. A., Vollmer, T. R., Jagtiani, S., Smith, R. G., & Mazaleski, J. L. (1993). Extinction of self-injurious escape behavior with and without instructional fading. Journal of Applied Behavior Analysis, 26, 353-360.