



Published in final edited form as:

*Behav Ther.* 2012 September ; 43(3): 679–685. doi:10.1016/j.beth.2011.10.001.

## Inferential Precision in Single-Case Time-Series Data Streams: How Well Does the EM Procedure Perform When Missing Observations Occur in Autocorrelated Data?

**Justin D. Smith,**

University of Tennessee Knoxville

**Jeffrey J. Borckardt,** and

Medical University of South Carolina

**Michael R. Nash**

University of Tennessee Knoxville

### Abstract

The case-based time-series design is a viable methodology for treatment outcome research. However, the literature has not fully addressed the problem of missing observations with such autocorrelated data streams. Mainly, to what extent do missing observations compromise inference when observations are not independent? Do the available missing data replacement procedures preserve inferential integrity? Does the extent of autocorrelation matter? We use Monte Carlo simulation modeling of a single-subject intervention study to address these questions. We find power sensitivity to be within acceptable limits across four proportions of missing observations (10%, 20%, 30%, and 40%) when missing data are replaced using the Expectation-Maximization Algorithm, more commonly known as the EM Procedure (Dempster, Laird, & Rubin, 1977). This applies to data streams with lag-1 autocorrelation estimates under 0.80. As autocorrelation estimates approach 0.80, the replacement procedure yields an unacceptable power profile. The implications of these findings and directions for future research are discussed.

### Keywords

autocorrelation; case-based; missing observations; power sensitivity; time-series

---

Experts have concluded that maximum likelihood methods are the preferred missing data management approach in group multivariate designs (e.g., Horton & Kleinman, 2007; Ibrahim, Chen, Lipsitz, & Herring, 2005; Ragnathan, 2004). Velicer and Colby (2005a, 2005b) extended this finding to time-series designs by establishing the superiority of maximum likelihood methods to other ad hoc procedures in the estimation of various critical time-series parameters. Velicer and Colby (2005a) applied Monte Carlo simulation to data streams with 100 observations per phase and autoregressive integrated moving averages analysis (ARIMA; Box & Jenkins, 1970) to test the accuracy of parameter estimates. They found the Expectation-Maximization Algorithm (more commonly known as the EM Procedure; Dempster, Laird, & Rubin, 1977) approach to be a good solution with minimal distortion. The EM Procedure is an algorithm derived from maximum likelihood estimation

techniques that is available in many common statistical software packages used in the social sciences. However, single-case time-series designs in intervention research rarely satisfy the minimum requirements for autoregressive integrated moving averages analysis used by Velicer and Colby ( $\approx 30\text{--}50$  observations per phase; e.g., Borckardt et al., 2008; Box & Jenkins, 1970). Thus, examination of the relevance of their conclusions in the context of more typical intervention research is warranted.

The time-series design is a longitudinal methodology that requires repeated observations over time (from a single unit or person) across conditions or phases. Among the social sciences, the time-series design occurs most naturally in the fields of physiological psychology, econometrics, and finance where each phase-of-interest has hundreds or thousands of observations that are tightly packed across time (e.g., electroencephalography actuarial data, financial market indices). However, time-series designs are now becoming more common in intervention research conducted in the clinical setting (e.g., Borckardt et al., 2008; Green, Rafaeli, Bolger, Shrout, & Reis, 2006; Kazdin, 2008; Smith, 2011).

Unlike the rich data streams generated by other disciplines, the typical clinical intervention data stream is much shorter, with baseline observation periods most commonly in the range of 8 to 12 measurement points (Center, Skiba, & Casey, 1986; Jones, Vaught, & Weinrott, 1977; Sharpley, 1987; Smith, 2011). Center et al. (1986) suggested 5 as the minimum number of baseline measurements in order to accurately estimate autocorrelation. However, this suggested benchmark doesn't always satisfy the requirements of time-series analytic methods, such as ARIMA. Longer baseline periods also increase the validity of an intervention effect and also reduce the biasing effects of autocorrelation (Huitema & McKean, 1994). Behavioral interventions targeting specific, and often singular, dependent variables are often much shorter, as evidenced by the findings of Huitema's (1985) review of 881 experiments published in the *Journal of Applied Behavior Analysis*, which resulted in a modal number of 3 to 4 baseline points. A number of recent time-series simulation studies have modeled their data after these very brief interventions (e.g., Manolov & Solanas, 2008; Manolov, Solanas, Sierra, & Evans, 2011; Solanas, Manolov, & Onghena, 2010). However, a recently completed systematic review of 409 recently published single-case experiments by Smith (2011) suggests that baseline periods have increased since the Huitema review of 1985: The Smith review found a mean number of 10.22 observations in the baseline phase. The present study addresses the limitations of these previous simulation studies in which modeling of time-series data streams occurred with either very short or very long baseline periods. This simulation study uses a 10 data-point baseline length, which is more likely to occur in applied clinical intervention research.

There are two real-world implications of conducting intervention research using a time-series design: Longitudinal designs of all types are inherently susceptible to missing observations (Laird, 1988), due in part to the fact that most research designs require observations to occur outside the controlled confines of a laboratory and across an extended period of time (weeks or months compared to seconds or minutes). Since time-series analysis techniques assume complete data, it is imperative that researchers identify a reliable method of substitution that maximizes the precision of our inferences regarding an intervention effect. Second, when we want to test intervention effects in short time-series data streams with single subjects, the paucity of observations compromises our ability to estimate effect above and beyond autocorrelation with few exceptions.

As time-series designs become more common in intervention research, it will be imperative to understand the way in which autocorrelation and missing values affect statistical inferences drawn from this type of data. Numerous statistical methods have been proposed in recent years to analyze single-subject research that accounts for autocorrelation (e.g.,

Borckardt et al., 2008; Manolov & Solanas, 2009; McKnight, McKean, & Huitema, 2000; Parker & Brossart, 2003; Parker & Vannest, 2009; Parker, Vannest, Davis, & Sauber, 2011; Solanas et al., 2010); yet, there appears to be no consensus regarding the best available method (e.g., Campbell, 2004; Kratochwill et al., 2010; Manolov et al., 2011; Parker & Brossart).

This Monte Carlo simulation study examines the impact of employing the EM Procedure for missing observations on time-series data streams typical of the case-based designs encountered in the clinical outcome literature (e.g., Borckardt et al., 2008; Kazdin, 1992; Lambert & Ogles, 2004; Smith, 2011). In contrast to the studies conducted by Velicer and Colby (2005a, 2005b), Manolov and Solanas (2008), Ferron and Sentovich (2002), and others, we are not necessarily concerned with the *accuracy* of maximum likelihood estimations (e.g., the EM Procedure) for missing observations in the context of model parameter fit. Instead, we focused on the precision of inferred significance based on time-series level-change (phase-effect) analyses when missing observations have been replaced using the EM Procedure. As the Velicer and Colby (2005a) study demonstrated, the EM Procedure results in relatively accurate estimations of the phase-effect statistics compared to other methods. However, to our knowledge there has yet to be an examination of the way in which these estimates are related to inferential precision. Based on conventional wisdom and previous research findings, we hypothesized that as the proportion of missing data increased, power sensitivity would deteriorate. Similarly, we hypothesized that power sensitivity would deteriorate as autocorrelation increased: Previous single-case time-series simulation studies have established the relationship between higher autocorrelation estimates and distorted effect-size estimations using various analytic techniques (e.g., Manolov & Solanas, 2008, 2009; Manolov et al., 2011), suggesting that autocorrelation will have an effect on inferential precision as well.

## Methods

### Outline of Study Procedures

1. Monte Carlo simulation was used to create 5,000 time-series data streams modeled after a typical 8-week study of psychotherapy effectiveness: 10 baseline and 56 intervention data points. Equal proportions of 5 levels of autocorrelation were programmed and were stable across both phases (0.0, 0.2, 0.4, 0.6, 0.8.)
2. Data points were removed completely at random from the data streams in four proportions: 10%, 20%, 30%, and 40%.
3. Missing observations were then replaced using the EM Procedure.
4. A simple effect of phase was calculated for each data stream using Glass's  $\Delta$ .
5. Power sensitivity was determined across each proportion of missing data and level of autocorrelation.

**The Monte Carlo Simulation of Data Streams**—In order to mimic the basic A-B design commonly used in treatment outcome research, our simulated data streams are comprised of a 10-day baseline phase and a 56-day (8-week) intervention period, with a daily measurement frequency (i.e., 1 day = 1 measurement). The majority of psychotherapy research is conducted on treatments offered once a week for not more than 14 weeks (Lambert, Bergin, & Garfield, 2003). We chose this simulated intervention period length in order to better simulate a typical psychotherapy treatment and also address the limitations in the literature of modeling simulation studies after very brief behavioral interventions (e.g., Manolov & Solanas, 2008; Manolov et al., 2011; Solanas et al., 2010). The longer data

streams also allowed for more accurate estimates of autocorrelation, particularly during the baseline period.

This study used a Monte Carlo simulation to create the time-series data. The simulations were conducted using custom-developed software written using RealBasic (RealSoftware Inc., 2011) by the second author.<sup>1</sup> The formula used for the simulations is presented as Equation 1.

$$y_i = \alpha \times S(n_1)_i + \varepsilon_i \quad (1)$$

Where:

- $\varepsilon_i$   $\rho\varepsilon_{i-1} + \varepsilon_i$
- $\rho$  autocorrelation
- $\alpha$  step size
- $\varepsilon_i$  indicates each is independently identically standard normally distributed:  $N(0,1)$
- $n_1$  Baseline phase length
- $n$  data stream length
- $i$  time

$S$  is a step function being 0 at times  $i=1, 2, 3 \dots n_1$  and 1 at times  $i=n_1+1, n_1+2, n_1+3 \dots n$ .

Manolov and Solanas (2008) determined that calculated effect sizes were linearly related to the autocorrelation of the data stream. Thus, 5,000 data streams were created with equal proportions of five different levels of programmed lag-1 autocorrelation that was stable across the two phases (0.0, 0.2, 0.4, 0.6, and 0.8), and a distribution of a mean effect size of 2 (Glass's  $\Delta$ ; [Mean of Phase B – Mean of Phase A] / Standard Deviation of Phase A), programmed to contain a set of random normal deviates with a standard deviation of 1 across each level of autocorrelation. The starting value of the step and the intercept was set to 0. The next 9 values were calculated using Equation 1, with 0 as a constant intercept before the step (i.e., the shift between the baseline and intervention phases). Since the simulated data streams contain a set of random normal deviates, the step size represents the effect size in question, which is the number of standard deviation for each series of simulations. The programmed autocorrelation values are not altered around the step and remain constant between both phases in the data stream. The described characteristics of the 5,000 data streams are prior to the removal of observations and subsequent substitution of missing values described in the remainder of this article.

**Removal of Observations**—One assumption of maximum likelihood estimation techniques, such as the EM Procedure, is that the data are either missing at random (MAR) or are missing completely at random (MCAR). Ibrahim et al. (2005) discussed these different classifications of missing data.<sup>2</sup> A third classification of missing values is those that are nonignorable or systematically missing from the data (e.g., the observation that is missing is a function of the value that is not being reported). For example, a suicidal patient may not report the severity of suicidal ideation on days in which these values would be very high. The current study removed data completely at random, satisfying the MCAR assumption. From the original data set of 5,000 data streams, four proportions of observations were randomly removed (10%, 20%, 30%, and 40%). These proportions are

<sup>1</sup>The algorithm used for the Monte Carlo simulation is available from the authors upon request.

<sup>2</sup>It should be noted that the standard definitions of MAR and MCAR pertain to the measurement of independent observations and not autocorrelated data.

identical to those of the Velicer and Colby (2005a) study, which was based on the findings of Rankin and Marsh's (1985) study of the effects of missing data on model fit.

**Replacing Missing Observations**—In a similar Monte Carlo simulation study, Velicer and Colby (2005a) demonstrated that maximum likelihood methods, such as the EM Procedure, yielded accurate estimates of time-series model parameters, but with significantly longer data streams than this and other time-series simulation studies. Horton and Kleinman (2007) report that for each missing observation, “multiple entries are created in an augmented dataset for each possible value of the missing covariates, and a probability of observing that value is estimated given the observed data and current parameter estimates” (p. 82). Rubin (1976) provides the justification that this is the correct likelihood to be maximized when data are MAR or MCAR. Furthermore, Shumway and Stoffer (1982) discussed the underlying processes of the EM Procedure when applied to time-series data streams. The EM Procedure was applied to all data streams in the data set using PASW Statistics for Windows and Mac. (2010) (Version 18.0.0). Chicago: SPSS Inc.

**Calculating an Effect of Phase**—Empirical studies (e.g., Ferron, Foster-Johnson, & Kromney, 2003; Ferron & Ware, 1995) suggest that Type II error is the main problem in single-case time-series data streams, since Type I error can usually be controlled using randomization elements in the research design of the study (for examples, see Ferron & Ware; Kratochwill & Levin, 2010). Therefore, power sensitivity is an appropriate statistic for determining precision because it is complementary to Type II error rates. Manolov and Solanas (2008) found that autocorrelation least affected effect sizes using standardized mean differences, such as Glass's  $\Delta$ , Cohen's  $d$ , and Hedge's  $g$ , and percentage of nonoverlapping data indices. Similarly, Solanas et al. (2010) concluded that level change effect sizes, calculated using a standardized mean differences approach, were “not affected by the presence of the degree of autocorrelation tested, as the values of its estimators do not change regardless of the values of  $\phi_1$  for AR(1) processes and  $\theta_1$  for MA(1) processes” (p. 536). Thus, following the lead of Ferron and Sentovich (2002), Manolov and Solanas (2008), Solanas et al., and others, we calculated an effect of phase using Glass's  $\Delta$ .<sup>3</sup> Using a standard critical values table, we determined the effect size needed to infer significance (two-tailed) with our data streams, using an alpha ( $\alpha$ ) of 0.05, was a Glass's  $\Delta$  of 1.023, based on 17 degrees of freedom (i.e., 16 estimatable lags of autocorrelation+the intercept; see Box & Jenkins, 1970).

**Determining Power Sensitivity**—Power sensitivity is the ability to detect an effect, if the effect actually exists (See Equation 2; Cohen). The acceptable rate of power sensitivity has been set traditionally at 0.80 (Cohen, 1988). Power sensitivity is influenced by the statistic used (e.g., Glass's  $\Delta$ ,  $r$ ,  $d$ ), the  $N$ ,  $\alpha$ , and the size of the effect. In this study, the  $N$  is relatively large with 66 observations in each data stream. We chose  $\alpha = 0.05$  because of its commonality, and we chose Glass's  $\Delta$  based on previous research.

$$\text{Power Sensitivity} = \frac{\text{number of True Positives}}{\text{number of True Positives} + \text{number of False Negatives}} \quad (2)$$

<sup>3</sup>We chose this method for a number of reasons even though there are many other potential methods for analyzing single-case time-series data: A mean differences approach does not account for autocorrelation in the data stream, and thus the relationship between autocorrelation and other parameters under investigation is not due to the way in which the analytic method used manages autocorrelation. Examining the inferential precision of different statistical analysis methods under the conditions of the data in this study is certainly an area for future research given the prevalence of available methods and the ongoing debate regarding their accuracy and utility. We certainly are not advocating the use of one statistical method over another. Researchers need to select the appropriate analytical method for their research design and the topic under investigation.

After Glass's  $\Delta$  was calculated for each individual data stream across all proportions of missing values, we used a dummy coding scheme to code significant or insignificant effects in the original data streams with no missing observations and those data streams where missing observations were replaced using the EM Procedure. This was done as a means of categorizing results in order to determine the power sensitivity across proportions of missing values and levels of lag 1 autocorrelation.

## Results

In order to address our first hypothesis, we calculated the power sensitivity for each proportion of missing observations, regardless of autocorrelation. The power sensitivity at each proportion of missing data was: 10%= .914, 20%= .913, 30%=.912, 40%=.910. These results indicate relative consistency across the four proportions of missing values and they are above 0.80, suggesting acceptable power sensitivity. To address our second hypothesis, we calculated the power sensitivity of each level of autocorrelation and proportion of missing observations. Figure 1 illustrates that the pattern of power sensitivity is similar across the four proportions of missing observations and deteriorates as autocorrelation increases, falling just below the acceptable level when autocorrelation reached 0.80.

## Conclusions

This study provides further support for the accuracy of the EM Procedure on the level-change parameter and also extends these findings to inform decision-making in research employing time-series designs. Our results suggest that the proportion of missing observations is not as significant a factor in inferential precision as the level of autocorrelation in the data stream: The EM Procedure seems to be an effective means of replacing missing observations and maintaining power sensitivity, even when 40% of the data stream is missing completely at random. However, when autocorrelation is in the vicinity of 0.80, regardless of the proportion of missing data, power sensitivity falls below the acceptable 0.80 level. Yet, overall power sensitivity remained quite high in our simulated sample, indicating that the EM Procedure is well-suited to managing missing observations in time-series data-streams. It is imperative to appreciate that these results only apply to missing data that is MCAR. Data that is MAR or systematically missing may not perform similarly. Additional research in this area is needed.

## Limitations and Future Directions for Research

The findings of this study occurred under nearly optimal conditions, unlike real-world applied research in which the assumption of MCAR is rarely satisfied, and the researcher doesn't have the privilege of "knowing the real effect size" without missing observations. A replication of these studies with naturally occurring data streams is the next logical step in determining the interaction between, and effects of, different time-series parameters on the conclusions drawn from time-series research. Alternatively, employing missing data algorithms with systematically missing observations, which are more likely to occur in applied research, would be useful for researchers conducting single-case time-series studies, as would comparing the accuracy of other available missing-data-estimation techniques, such as multiple imputation and Bayesian.

Second, this study was modeled after an 8-week treatment outcome study with a 10-day pretreatment baseline, which limits generalizability to longer or shorter time periods. Manolov and Solanas (2008) did not find differences due to the length of the data stream, but their streams were rather short compared to the single-subject psychotherapy outcome experiment after which our study was modeled. Given the results of previous simulation studies (e.g., Manolov & Solanas, 2008; Velicer & Colby, 2005a), we believe that our

primary findings will hold true in studies of longer and shorter data streams, but this is an area of inquiry that still requires controlled investigation. Future studies with various data stream lengths that might be encountered in applied research settings would improve generalizability.

Last, this study only employed one statistical approach for detecting an effect of phase. However, there are many available options for the analysis of single-case time-series experiments, many of which would arguably be more appropriate in applied research with autocorrelated data. The available statistical methods designed for time-series analysis account for the autocorrelation of the data stream differently, and calculate effect sizes differently as well, which has resulted in a great deal of debate in the field (e.g., Campbell, 2004; Cohen, 1994; Ferron & Sentovich, 2002; Ferron & Ware, 1995; Kirk, 1996; Manolov & Solanas, 2008; Olive & Smith, 2005; Parker & Brossart, 2003; Robey, Schultz, Crawford, & Sinner, 1999; Velicer & Fava, 2003). Future simulation studies might consider comparing inferential precision of multiple statistical analysis methods, preferably those found to be superior in previous research (e.g., Campbell, 2004; Manolov et al., 2011; Parker & Brossart, 2003).

## Acknowledgments

Research support for the first author was provided by institutional research training grant MH20012 from the National Institute of Mental Health awarded to Elizabeth A. Stormshak.

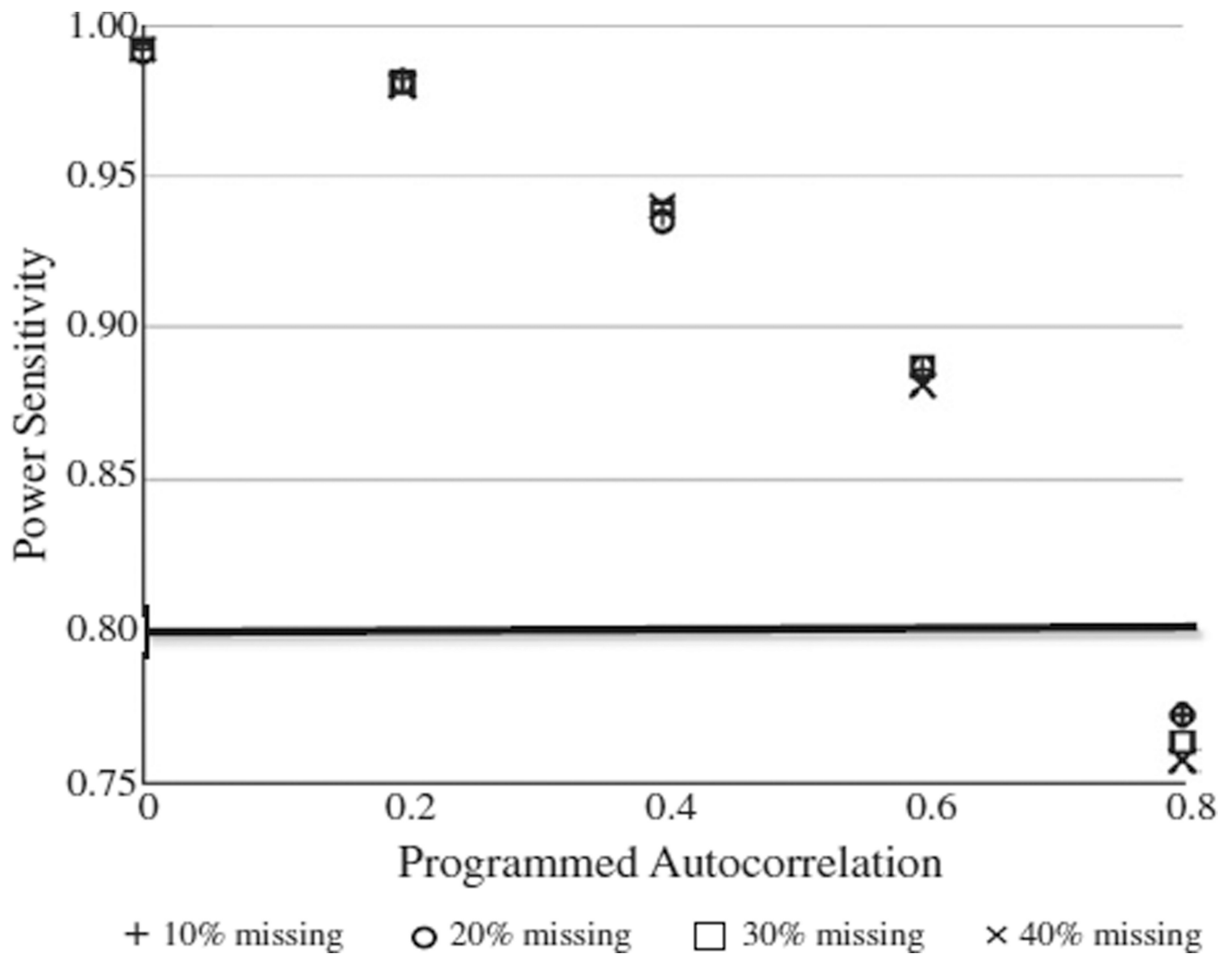
## References

- Borckardt JJ, Nash MR, Murphy MD, Moore M, Shaw D, O'Neil P. Clinical practice as natural laboratory for psychotherapy research. *American Psychologist*. 2008; 63:1–19. [PubMed: 18193977]
- Box, GEP.; Jenkins, GM. *Time-series analysis: Forecasting and control*. San Francisco: Holden-Day; 1970.
- Campbell JM. Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*. 2004; 28:234–246. [PubMed: 14997950]
- Center BA, Skiba RJ, Casey A. A methodology for the quantitative synthesis of intra-subject design research. *Journal of Educational Science*. 1986; 19:387–400.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
- Cohen J. The earth is round (p .05). *American Psychologist*. 1994; 49:997–1003.
- Dempster A, Laird N, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*. 1977; 39(1):1–38.
- Ferron J, Foster-Johnson L, Kromney JD. The functioning of single-case randomization tests with and without random assignment. *The Journal of Experimental Education*. 2003; 71:267–288.
- Ferron J, Sentovich C. Statistical power of randomization tests used with multiple-baseline designs. *The Journal of Experimental Education*. 2002; 70:165–178.
- Ferron J, Ware W. Analyzing single-case data: The power of randomization tests. *The Journal of Experimental Education*. 1995; 63:167–178.
- Green AS, Rafaeli E, Bolger N, Shrout PE, Reis HT. Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*. 2006; 11:87–105. [PubMed: 16594769]
- Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*. 2007; 61:79–90. [PubMed: 17401454]
- Huitema BE. Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment*. 1985; 7:107–118.
- Huitema BE, McKean JW. Reduced bias autocorrelation estimation: Three jackknife methods. *Educational and Psychological Measurement*. 1994; 54:654–665.

- Ibrahim JG, Chen M-H, Lipsitz SR, Herring AH. Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*. 2005; 100(469):332–346.
- Jones RR, Vaught RS, Weinrott MR. Time-series analysis in operant research. *Journal of Behavior Analysis*. 1977; 10:151–166.
- Kazdin, AE. *Research designs in clinical psychology*. 2nd ed. Boston: Allyn & Bacon; 1992.
- Kazdin AE. Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*. 2008; 63:146–159. [PubMed: 18377105]
- Kirk RE. Practical significance: A concept whose time has come. *Educational and Psychological Measurement*. 1996; 56:746–759.
- Kratochwill, TR.; Hitchcock, J.; Horner, RH.; Levin, JR.; Odom, SL.; Rindskopf, DM.; Shadish, WR. Single-case designs technical documentation. 2010. Retrieved from The What Works Clearinghouse website: [http://ies.ed.gov/ncee/wwc/pdf/wwc\\_scd.pdf](http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf)
- Kratochwill TR, Levin JR. Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*. 2010; 15:124–144. [PubMed: 20515235]
- Laird N. Missing data in longitudinal studies. *Statistics in Medicine*. 1988; 7(1–2):305–315. [PubMed: 3353609]
- Lambert, MJ.; Bergin, AE.; Garfield, SL. Introduction and historical overview. In: Lambert, MJ., editor. *Bergin and Garfield's handbook of psychotherapy and behavior change*. New York: John Wiley & Sons, Inc.; 2003. p. 3-15.
- Lambert, MJ.; Ogles, BM. Lambert, MJ. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. 5th ed. New York: John Wiley & Sons; 2004. The efficacy and effectiveness of psychotherapy; p. 139-193.
- Manolov R, Solanas A. Comparing  $N=1$  effect sizes in presence of autocorrelation. *Behavior Modification*. 2008; 32:860–875. [PubMed: 18497425]
- Manolov R, Solanas A. Problems of the randomization test for AB designs. *Psicológica*. 2009; 30:137–154.
- Manolov R, Solanas A, Sierra V, Evans JJ. Choosing among techniques for quantifying single-case intervention effectiveness. *Behavior Therapy*. 2011; 42:533–545. [PubMed: 21658534]
- McKnight SD, McKean JW, Huitema BE. A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods*. 2000; 5:87–101. [PubMed: 10937324]
- Olive ML, Smith BW. Effect size calculations and single subject designs. *Educational Psychology*. 2005; 25(2–3):313–324.
- Parker RI, Brossart DF. Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy*. 2003; 34:189–211.
- Parker RI, Vannest K. An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*. 2009; 40:357–367. [PubMed: 19892081]
- Parker RI, Vannest K, Davis JL, Sauber SB. Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*. 2011; 42:284–299. [PubMed: 21496513]
- Version 18.0.0. Chicago: SPSS Inc.; 2010. PASW Statistics for Windows and Mac.
- Ragunathan TE. What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health*. 2004; 25:99–117.
- Rankin ED, Marsh JC. Effects of missing data on the statistical analysis of clinical time-series. *Social Work Research and Abstracts*. 1985; 21:13–16.
- RealSoftware Inc. *RealBasic 3*. Austin, TX: 2011.
- Robey RR, Schultz MC, Crawford AB, Sinner CA. Single-subject clinical-outcome research: Designs, data, effect sizes, and analyses. *Aphasiology*. 1999; 13:445–473.
- Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–590.
- Sharpley CF. Time-series analysis of behavioural data: An update. *Behaviour Change*. 1987; 4:40–45.
- Shumway RH, Stoffer DS. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time-Series Analysis*. 1982; 3(4):253–264.
- Smith JD. Single-Case Experimental Designs: A Systematic Review of Published Research and Current Standards. 2011 Manuscript submitted for publication.



- Solanas A, Manolov R, Onghena P. Estimating slope and level change in N = 1 designs. *Behavior Modification*. 2010; 34:195–218. [PubMed: 20234005]
- Velicer WF, Colby SM. A comparison of missing-data procedures for ARIMA time-series analysis. *Educational and Psychological Measurement*. 2005a; 65:596–615.
- Velicer, WF.; Colby, SM. Maydeu-Olivares, A.; McArdle, JJ. Contemporary psychometrics. A Festschrift to Roderick P. McDonald. Hillsdale, NJ: Lawrence Erlbaum; 2005b. Missing data and the general transformation approach to time series analysis; p. 509-535.
- Velicer, WF.; Fava, JL. Time series analysis. In: Schinka, J.; Velicer, WF., editors. *Research methods in psychology* (Vol. 2 of *Handbook of Psychology*, I. B. Weiner, Editor-in-Chief). New York: John Wiley & Sons; 2003.



**FIGURE 1.**  
*Power Sensitivity Across Proportions of Missing Observations at Each Level of Programmed Lag 1 Autocorrelation. Note. The bold horizontal line indicates Cohen's (1988) benchmark for power sensitivity (0.80).*