

Choosing among techniques for quantifying single-case intervention effectiveness

Rumen Manolov¹, Antonio Solanas^{1,4},
Vicenta Sierra², and Jonathan J. Evans³

¹ Department of Behavioral Sciences Methods, Faculty of Psychology,
University of Barcelona.

² Department of Quantitative Methods, ESADE-Ramon Llull University.

³ Centre for Population and Health Sciences, College of Medicine, Veterinary
and Life Sciences, University of Glasgow.

⁴ Institute for Research in Brain, Cognition, and Behavior (IR3C).

MAILING ADDRESS

Correspondence concerning this article should be addressed to Rumen Manolov, Departament de Metodologia de les Ciències del Comportament, Facultat de Psicologia, Universitat de Barcelona, Passeig de la Vall d'Hebron, 171, 08035-Barcelona, Spain. Phone number: +34934031137. Fax: +34934021359. Electronic mail may be sent to Rumen Manolov at rmenov13@ub.edu.

AUTHORS' NOTE

This research was partially supported by the Spanish Ministry of Science and Innovation, grant *PSI2009-07076*, and by the Generalitat of Catalonia's Ministry of Universities, Research and the Information Society grant *2009SGR1492*.

RUNNING HEAD

Quantifying intervention effectiveness

ABSTRACT

If single case experimental designs are to be used to establish guidelines for evidence-based interventions in clinical and educational settings, numerical values that reflect treatment effect sizes are required. The present study compares four recently developed procedures for quantifying the magnitude of intervention effect using data with known characteristics. Monte Carlo methods were used to generate AB designs data with potential confounding variables (serial dependence, linear and curvilinear trend, and heteroscedasticity between phases) and two types of treatment effect (level and slope change). The results suggest that data features are important for choosing the appropriate procedure and, thus, inspecting the graphed data visually is a necessary initial stage. In the presence of serial dependence or a change in data variability, the Nonoverlap of All Pairs (NAP) and the Slope and Level Change (SLC) were the only techniques of the four examined that performed adequately. Introducing a data correction step in NAP renders it unaffected by linear trend, as is also the case for the Percentage of Nonoverlapping Corrected Data and SLC. The performance of these techniques indicates that professionals' judgments concerning treatment effectiveness can be readily complemented by both visual and statistical analyses. A flowchart to guide selection of techniques according to the data characteristics identified by visual inspection is provided.

Key words: single-case, effect size, autocorrelation, trend

Single-case experimental designs (SCEDs) have been shown to be useful for evaluating intervention effectiveness in several behavioral fields (Blampied, 2000), including educational (Horner, Carr, Halle, McGee, Odom, & Wolery, 2005; Parker & Brossart, 2006) and clinical psychology settings (Callahan & Barisa, 2005; Perdices & Tate, 2010). Evidence-based guidelines in relation to treatment interventions can be established by collating data from multiple SCEDs (Kratochwill & Levin, 2010). Some form of measure is needed to summarize the results of the study, not only to meet a quality criterion for N=1 research (Horner et al., 2005), but also for accountability (Chambless & Ollendick, 2001), communication between researchers, and particularly to enable conducting meta-analyses (Busse, Kratochwill, & Elliott, 1995). The latter are fundamental to evidence-based practice, given that the clinician who wants to select the correct treatment is interested in syntheses (i.e., meta-analysis) of data rather than in individual studies (Kratochwill, 2007). Although the call for evidence-based practice has emphasized the importance of summary measures (Shadish, Rindskopf, & Hedges, 2008), one should not lose sight of the fact SCED studies provide important information on individual participant characteristics, time course of interventions, setting, and behavior measured (Reichardt, 2006) that may address the question of what treatments are useful in particular situations (Chambless & Ollendick, 2001).

Single-case investigation and statistical reasoning are clearly not incompatible (White, Rusch, Kazdin, & Hartmann, 1989). In fact, over the last decades a considerable number of methods of analyzing SCED data have been

proposed (Allison & Gorman, 1993; Borckardt, Nash, Murphy, Moore, Shaw, & O'Neil, 2008; Center, Skiba, & Casey, 1985-1986; Ma, 2006; Manolov & Solanas, 2009; Parker, Hagan-Burke, & Vannest, 2007; Parker & Vannest, 2009; Parker, Vannest, & Brown, 2009; Scruggs, Mastropieri, & Casto, 1987; Solanas, Manolov, & Onghena, 2010). Despite the proliferation of methods, there is currently no clear consensus on which is the most appropriate one for analyzing SCED data. For the present study four recently developed techniques were selected and compared: Improvement Rate Difference (IRD; Parker et al., 2009), Nonoverlap of All Pairs (NAP; Parker & Vannest, 2009), Percentage of Nonoverlapping Corrected Data (PNCD; Manolov & Solanas, 2009), and Slope and Level Change (SLC; Solanas et al., 2010). There were several reasons for the selection of these techniques. Firstly, they have not yet been subjected to extensive testing in order to establish their appropriateness for a variety of data. Some of the techniques are modifications of previously presented methods and were designed to improve upon them. For example, PNCD was selected instead of the Percent of Nonoverlapping Data (Scruggs et al., 1987) as a previous study had shown PNCD to be superior to PND (Manolov & Solanas, 2009). Another related method, the Percentage of data points Exceeding the Median (Ma, 2006) was excluded due to poor results (Parker & Hagan-Burke, 2007; Wolery, Busick, Reichow, & Barton, 2010). IRD was chosen instead of its related original method, Percentage of All Nonoverlapping Data (Parker et al., 2007), since it has been argued that the latter is too complicated for routine use (Parker & Vannest, 2009; Schneider,

Goldstein, & Parker, 2008). The second reason for choosing IRD, NAP, PNCD, and SLC is that they share two features that make them potentially attractive to and useful for applied researchers – ease of computation and ease of interpretation of their values (see Appendix 1 for details). Straightforward techniques are preferred since they require less effort from professionals and are less likely to be used inappropriately. Finally, the four techniques selected can readily be used to complement visual inspection, a synergy that has repeatedly been called for (Fisch, 2001; Parker, Cryer, & Byrns, 2006).

The objective of the present study is to extend the existing evidence on the performance of IRD, NAP, PNCD, and SLC and compare them directly in terms of a) the degree to which they are affected by potential confounding variables (i.e., serial dependence, heteroscedasticity, and trend) and b) the degree to which they detect different types and magnitudes of treatment effect. Thus, the context of comparison of the four approaches to analysis of SCED data is simulated data with known characteristics. IRD and NAP have previously been tested with real behavioral data in terms of typical values, discriminability, and relatedness to other effect size methods (Parker & Vannest, 2009; Parker et al., 2009) and the present study intends to complement this information. As regards PNCD and SLC, for which simulation studies have already been carried out (Manolov & Solanas, 2009, and Solanas et al., 2010, respectively), the novelty consists in introducing two previously unstudied potential confounding variables – heteroscedasticity and curvilinear quadratic trend.

Although both new data features studied are conceptualized here as confounding variables, some comments are necessary. Regarding variability, it should be pointed out that in some cases the aim of a treatment may be to decrease the variability of client's behavior. As regards quadratic trend (i.e., a progressive change with an increasing rate), according to the moment in which the intervention is introduced such a curve can have different interpretations. For instance, if the situation is like the one depicted on the left panel of Figure 1, visual inspection may suggest that the data represent a change in slope or a delayed treatment effect. For the case shown on the right panel of Figure 1 it seems clearer that the behavior is experiencing a trend not related to the treatment and, therefore, the assessment of intervention effectiveness would have a different outcome. Interpretations become more difficult if the intervention point is approximately in the middle of a similar data pattern.

INSERT FIGURE 1 ABOUT HERE

Thus the main aims are: a) to determine whether autocorrelation, unequal variances, and trends distort the magnitude of effect quantifications and to what extent; and b) to test which of the techniques is most sensitive to several degrees of change in level and change in slope in the data. A secondary goal is to test the relevance of phase length, data generation processes, and random variable distributions for both distortion and effect detection.

Method

Design and series lengths studied

The procedures are tested in the context of the AB design (a baseline phase followed by treatment introduction), but the conditions also represent pairs of adjacent AB phases within a more complex structure (e.g., multiple-baseline and ABAB). Relatively short data series ($n = n_A + n_B$) were included in the present study, reflecting the typical length of data collection periods in applied settings: a) $n = 10$ with $n_A = n_B = 5$; b) $n = 20$ with $n_A = n_B = 10$; and c) $n = 30$ with $n_A = n_B = 15$.

Data generation

Monte Carlo methods were used to generate data with different types of effects with different magnitudes, as well as potential confounding factors in order to compare these known features to the values yielded by the techniques. To the best of our knowledge, heteroscedasticity and curvilinear trend have not previously been simulated in studies on magnitude of effect techniques.

The data generation model used was (Huitema & McKean, 2000):

$$y_t = \beta_0 + \beta_1 \cdot T_t + \beta_2 \cdot LC_t + \beta_3 \cdot SC_t + \varepsilon_t,$$

where y_t is the dependent variable measurement at moment t , β_0 is intercept set to zero, β_1 , β_2 , and β_3 are the coefficients associated with trend, level change, and slope change, respectively, T_t is a dummy variable for trend at moment t taking values from 1 to n for linear trend and from 1^2 to n^2 for quadratic curvilinear trend, LC_t is a dummy variable for level change (0 for phase A and

to 1 for phase B), SC_t is a dummy variable for slope change (0 for phase A and taking values from 0 to n_B-1 for phase B), and ε_t is the error term.

Following the discussion by Harrop and Velicer (1985) on the most useful models to represent behavioral data, the error term (ε_t) was generated by a first-order autoregressive process, AR(1), $\varepsilon_t = \varphi_1 \cdot \varepsilon_{t-1} + u_t$ with φ_1 ranging from $-.3$ to $.6$ in steps of $.3$ and by a first-order moving average process, MA(1), $\varepsilon_t = u_t - \theta_1 \cdot u_{t-1}$ (McCleary & Hay, 1980) with θ_1 ranging from $-.9$ to $.3$ in steps of $.3$. In the latter case, the values of θ_1 correspond to φ_1 from $-.2752$ to $.4972$, since $\varphi_1 = -\theta_1/(1 + \theta_1^2)$.

For both processes three different distributions (exponential, normal, and uniform) were used for generating the random variable u_t in order to represent a wider range of real behavioral data (Bradley, 1977; Micceri, 1989) and, specifically, due to the differences among the distributions in terms of skewness and kurtosis. All distributions had a zero mean and a unity standard deviation for both phases for homoscedastic conditions. For the heteroscedastic conditions, two cases were specified – an increased (a standard deviation of 2) or decreased (a standard deviation of .5) phase B variability.

The β_1 , β_2 , and β_3 coefficients were set to vary from 0 to $.3$ in steps of $.1$ in order to permit studying the detection of different magnitudes of effect. For testing the effect of curvilinear trend β_1 was set to $.01$. The specific values were chosen by trial and error (Brossart, Parker, Olson, & Mahadevan, 2006; Parker & Brossart, 2003) to avoid floor and ceiling effects in the indices. It should be noted that a slope change of say $.2$ represents a greater effect than a

constant level change of .2 (more so for longer data series), since the former is progressive and thus increasing the average level with each measurement time.

Simulation

Fortran programs were used to carry out the following steps: 1) Vary systematically ϕ_1 or θ_1 . 2) Vary systematically β_1, β_2 , and β_3 . 3) Iterate 100,000 times steps 4 through 9. 4) Generate the u_t term according to a negative exponential, a normal, or a uniform distribution. 5) Set $\varepsilon_1 = u_1$. 6) Obtain the error term ε_t out of the random variable u_t using the AR(1) or the MA(1) model. 7) Obtain the dummy variables for trend, level and slope change. 8) Obtain the y_t array representing behavioral measurements. 9) Apply the four procedures to the generated data array y_t . 10) Average the obtained techniques' values from the 100,000 replications of each experimental condition.

For data generation NAG libraries *nag_rand_neg_exp*, *nag_rand_normal*, and *nag_rand_uniform* were used, making individual calls for each specific experimental condition. As a control against artificial effects (Greenwood & Matyas, 1990) and dependence between successive data series (Huitema, McKean, & McKnight, 1999), the first 50 data generated at each call to the libraries were eliminated in step 4.

Data analysis

The comparisons among the procedures were carried out in terms of the distorting effect of the confounding variables studied (i.e., autocorrelation,

heteroscedasticity, and general trend) and in terms of detection of the two types of intervention effect (i.e., level change and slope change). To quantify the amount of change in the indices associated with these data features ratios were computed, maintaining constant the features not included in the calculi:

a) for the degree of distortion due to autocorrelation, a ratio was computed between the magnitude of effect quantification for conditions with serially related data (in the numerator) and independent data (in the denominator);

b) for the degree of distortion due to unequal variability, a ratio was computed between the magnitude of effect quantification for heteroscedasticity (in the numerator) and homoscedasticity (in the denominator);

c) for the degree of distortion due to either type of trend, a ratio was computed between the magnitude of effect quantification for presence (in the numerator) and absence (in the denominator) of trend;

d) for the degree of detection of treatment effects, a ratio was computed between the magnitude of effect quantification for presence (in the numerator) and absence (in the denominator) of effect.

Regarding distortions, ratios close to 1 indicate that the confounding variable does not affect the value yielded by the procedure; a desirable characteristic. Ratios greater than 1 suggest that the magnitudes of effect are overestimated, whereas ratios smaller than 1 are indicative of underestimation. As regards effect detection, a ratio of 1 implies that an existing change in the measurements is missed by the procedure, which does not discriminate

between absence and presence of effect. Ratios greater than 1 are expected as indicators of the sensitivity to (different magnitudes of) treatment effects.

Results

Distortion due to autocorrelation

SLC is not affected by the presence of the degrees of autocorrelation tested, as the values of its estimators do not change regardless of the values of φ_1 for AR(1) processes and θ_1 for MA(1) processes. This result may be related to the data correction step, which has been shown to attenuate the affection by serial dependence in PNCD in comparison to PND (Manolov & Solanas, 2009), and considering that when a linear trend is present, data are positively serially related. However, these findings are restricted to the conditions tested – homogeneous AB data in which autocorrelation is constant in both phases. For low levels of positive and negative serial dependence the underestimation and overestimation of the remaining three indices is only slight, as shown on Tables 1 and 2. For $\varphi_1 \geq .5$ NAP maintains the desirable performance, whereas IRD and PNCD show greater distortion, with the latter being somewhat less affected for normal and uniform random variable (u_t) terms. These findings are common to both AR(1) and MA(1) processes. When treatment effect was simulated (results not shown here), the distortions occur in IRD and PNCD for the highest degrees of serial dependence tested.

INSERT TABLES 1 AND 2 ABOUT HERE

Distortion due to heteroscedasticity

NAP remains unaffected by heteroscedasticity, except for conditions with an exponential u_t term. SLC is also not affected by unequal variability, since the values of its estimators equal zero in the absence of treatment effect regardless of the variability in the series. Greater phase B variability results in an increase in the magnitude of effect calculated by PNCD and IRD, the latter being more affected (see Table 3). In contrast, lower phase B variability is only associated with higher values of IRD, while it leads to lower values for PNCD. The affection of IRD becomes more obvious for larger n .

INSERT TABLE 3 ABOUT HERE

The decrease in the variability of a behavior of interest after an intervention is potentially an important and desirable result and an index sensitive to such an effect might be useful for practitioners. This is not the case for NAP or SLC. The results for IRD and PNCD also suggest that neither of the procedures is suitable. PNCD marks the reduced variability with lower percentages, which may be misleading as such values are also indicative of a lack of effect. IRD yields greater values for both increased and decreased data variability and, thus, is not useful for discriminating between these conditions.

Distortion due to trend

PNCD and SLC are practically unaffected by the presence of linear trend in data, which indicates that the data correction step is effective for data exhibiting such a pattern. The values of SLC's slope and level change estimators are not presented in the tables, since they are equal to zero (for $\beta_2 = \beta_3 = 0$), regardless of the value of β_1 . For NAP, and to a greater extent for IRD, linear trend is associated with an inflation of the magnitude of effect quantifications, especially for exponential u_t term (Table 4). Thus, considered alone the values from these procedures may lead to an incorrect conclusion that an intervention is effective. Curvilinear trend affects all four procedures, particularly IRD (see Table 5). In the case of PNCD and SLC, these results suggest that the data correction step is only useful for linear trends. For both types of trend, the distortion in the procedures affected is greater for longer data series.

INSERT TABLES 4 AND 5 ABOUT HERE

Simulated treatment effect detection

Prior to discussing the results, it should be noted that for NAP ratios can only become as large as 2, since 1 is the maximum value of the index and .5 is the value expected (and actually obtained) for absence of effect.

IRD is the method that shows greater relative increment for both level change (Table 6) and slope change quantifications (Table 7). NAP and PNCD

perform similarly detecting level change, whereas for slope change the latter is more sensitive. For SLC the results showed that the average slope change estimate is equal to β_3 and the average level change estimate is equal to β_2 , an indication of the lack of bias. Ratios were not computed, since the estimate obtained by either estimator in the absence of an effect ($\beta_2 = \beta_3 = 0$) is zero.

INSERT TABLES 6 AND 7 ABOUT HERE

For level change, NAP yields approximately the same values irrespective of n , in contrast with the results for IRD, which discriminates better between presence and absence of effect for longer data series. For this kind of effect IRD shows greater sensitivity for exponential u_t terms. Slope change was better detected by all three procedures, which is emphasized for greater sample sizes, as progressive effects become more evident with each measurement. Another result to be highlighted in relation to series length is that the values yielded by PNCD and IRD become smaller as n increases, *ceteris paribus*.

Discussion

The present study compared four procedures for quantifying effect size in N=1 designs in order to help applied researchers choose among them when assessing intervention effectiveness. In the following section the main results will be summarized and used to make recommendations to practitioners.

Afterwards, the place of quantitative procedures in professionals' decision making process will be discussed. Finally, the limitations of the study will be pointed out alongside questions that need further investigation.

Performance of the techniques studied

The somewhat simplified summary presented in Table 8 shows that the effect size quantifications provided by some techniques (i.e., IRD) are distorted by all of the confounding variables studied, whereas others (NAP and SLC) are affected only by trend. Taking into account the results on NAP, the introduction of an initial data correction step identical to SLC's first step is proposed in order to deal with linear upward or downward trend. Simulations support this proposal (referred to as NAPcorr) and R code is presented in Appendix 2 for carrying out the necessary computations. The information on distortion needs to be complemented considering the sensitivity of the procedures which is higher for IRD. Among the remaining techniques, PNCD is more sensitive than NAP to slope changes. Although not directly comparable, SLC detects the effects and estimates them precisely.

INSERT TABLE 8 ABOUT HERE

Focusing on IRD, a limitation is that there is not a single approach to computing the index, since when k data points need to be eliminated in order to reach a complete nonoverlap, there may be several possible ways of

selecting k data points across phases A and B. Parker and his colleagues (2009) recommend balancing the amount of data points to eliminate from each phase. While performing the present study it was found that the solution in which $k/2$ values are eliminated from each phase leads to a median value of IRD compared to eliminating more data points from either phase. Hence, Parker et al.'s recommendation would lead to avoiding extreme (i.e., too low or too high) values of the magnitude of effect. It should be pointed out that when $n_A = n_B$, the specific bipartitioning of the k necessary eliminations does not affect the value of IRD. These considerations, in conjunction with the distortion by confounding variables, suggest that among the percentage indices, it is more advisable to use NAP and PNCD than IRD.

Concerning the secondary goals of the study, generating data from AR(1) or MA(1) processes with either of the random variable distributions did neither alter the general findings on the degree of distortion or sensitivity, nor the ranking among the most affected or sensitive procedures. Nonetheless, some particularities for the u_t terms were pointed out and the dependence of IRD and PNCD on series length was underlined.

Intervention effectiveness assessment

When assessing intervention effectiveness, quantification only complements the substantive criteria related to the field of expertise. Practical significance can be assessed considering both social and clinical clues, such as the perceptions of the client and of significant others (Kratochwill & Levin, 2010)

and moving from dysfunctional to functional ranges of behavior (Perdices & Tate, 2010), considering a variety of factors and indicators (Kazdin, 1999).

In order to improve objectivity and results documentation, suitable complements to psychologists' judgments have to be identified. Reviews suggest that visual inspection is most frequently applied (Parker & Brossart, 2003), although in some areas of psychology it is common to combine both statistical and visual analysis (Perdices & Tate, 2010). Visual inspection allows considering simultaneously several data attributes such as the immediacy of treatment effects, the presence of trends, and data variability within and between phases (Horner et al., 2005; Parker et al., 2006). Furthermore, plotting each measurement as it is obtained allows understanding better the behavioral processes and interrupting immediately undesired situations (Fahmie & Hanley, 2008). The whole data series can also be assessed visually in order to judge whether the data pattern corresponds to the expected one considering the design structure, which would be indicative of experimental control. Finally, the results presented here suggest that visual inspection may aid choosing the appropriate quantitative procedure.

Nonetheless, visual inspection used on its own does not provide formal decision rules. Therefore, the third pillar of effectiveness assessment (apart from substantive knowledge and visual inspection) should be a quantification of the strength of association between intervention and behavior. In the following section we provide guidance on selecting the appropriate technique and highlight the importance of initial visual inspection in that process.

Choosing among the techniques studied

On the basis of the results on distortion due to confounding variables and sensitivity to intervention effects, a quantitative procedure can be selected according to the data features identified by visual analysis (see Figure 2). In the case of autocorrelation, judgments need to be based on theoretical or data collection considerations (e.g., the interval between measurements), since estimation is biased in short series (Huitema & McKean, 1991).

INSERT FIGURE 2 ABOUT HERE

Several explanations are needed regarding the flowchart. For quadratic trend, visual inspection has to reveal where the greatest progressive change in the behavior starts. If a quadratic trend is identified during baseline, the practitioner would need to postpone the intervention until greater stability in the behavior measured is reached. Conversely, when the data pattern resembles a change in slope or a delayed effect is present, in which case the professional's familiarity with the client is critical. This in-depth knowledge is crucial, since the detection of delayed effects has been shown to be troublesome with fictitious data (Lieberman, Yoder, Reichow, & Wolery, 2010). For autocorrelated and/or heteroscedastic data NAP and SLC can be employed given that they provide different information. The same is the case for NAPcorr and SLC when a linear trend in the measurements is identified.

Limitations and future research

In the present simulation the curvilinear trend simulated was quadratic, since in short data series it is not likely to clearly identify higher order trends (e.g., a cubic trend implies a pattern with two marked apices). Outliers were not incorporated in the data generation model and future studies need to explore their effect. If outliers prove to have an impact on statistical procedures, visual analysis would be appropriate for detecting extreme values, while expert's knowledge on the particular case would be necessary for interpreting them. It would also be interesting to investigate the performance of the techniques in case of a different degree of autocorrelation across phases. There is evidence in some fields of this phenomenon (e.g., Robey, Schultz, Crawford, & Sinner, 1999), and, in fact, it would be logical to expect greater autocorrelation in phase B in cases of upward or downward trends (i.e., a progressive change in the behavior which makes it more homogeneous). Finally, the recommendation to use NAP and SLC is based on the experimental conditions studied and it does not exclude the utility of other techniques not tested here.

As no single procedure is superior in all conditions and the choice of technique depends on the nature of the data (e.g., presence of trend), the comparability between studies using different magnitude of effect measures is compromised and this represents an obstacle for performing meta-analyses. A possible solution is to combine studies in terms of small/medium/large effect sizes. Conventions on the meaning of each of these labels in a specific field and in relation to each statistical technique therefore need to be established.

References

- Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy*, *31*, 621-631.
- Blampied, N. M. (2000). Single-case research designs: A neglected alternative. *American Psychologist*, *55*, 960.
- Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. *American Psychologist*, *63*, 77-95.
- Bradley, J. V. (1977). A common situation conducive to bizarre distribution shapes. *American Statistician*, *31*, 147-150.
- Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification*, *30*, 531-563.
- Busse, R. T., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology*, *33*, 269-285.
- Callahan, C. D., & Barisa, M. T. (2005). Statistical process control and rehabilitation outcome: The single-subject design reconsidered. *Rehabilitation Psychology*, *50*, 24-33.

- Center, B. A., Skiba, R. J., & Casey, A. (1985-1986). A methodology for the quantitative synthesis of intra-subject design research. *The Journal of Special Education, 19*, 387-400.
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685-716.
- Fahmie, T. A., & Hanley, G. P. (2008). Progressing toward data intimacy: A review of within-session data analysis. *Journal of Applied Behavior Analysis, 41*, 319-331.
- Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models? *Behavioural Processes, 54*, 137-154.
- Greenwood, K. M., & Matyas, T. A. (1990). Problems with application of interrupted time series analysis for brief single-subject data. *Behavioral Assessment, 12*, 355-370.
- Harrop, J. W., & Velicer, W. F. (1985). A comparison of alternative approaches to the analysis of interrupted time-series. *Multivariate Behavioral Research, 20*, 27-44.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165-179.
- Huitema, B. E., & McKean, J. W. (1991). Autocorrelation estimation and inference with small samples. *Psychological Bulletin, 110*, 291-304.

- Huitema, B. E., & McKean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement, 60*, 38-58.
- Huitema, B. E., McKean, J. W., & McKnight, S. (1999). Autocorrelation effects on least-squares intervention analysis of short time series. *Educational and Psychological Measurement, 59*, 767-786.
- Kazdin, A. E. (1999). The meanings and measurements of clinical significance. *Journal of Consulting and Clinical Psychology, 67*, 332-339.
- Kratochwill, T. R. (2007). Preparing psychologists for evidence-based school practice: Lessons learned and challenges ahead. *American Psychologist, 62*, 829-843.
- Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15*, 124-144.
- Lieberman, R. G., Yoder, P. J., Reichow, B., & Wolery, M. (2010). Visual analysis of multiple baseline across participants graphs when change is delayed. *School Psychology Quarterly, 25*, 28-44.
- Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject research: Percentage of data points exceeding the median. *Behavior Modification, 30*, 598-617.
- Manolov, R., & Solanas, A. (2009). Percentage of nonoverlapping corrected data. *Behavior Research Methods, 41*, 1262-1271.

- McCleary, R., & Hay, R. A., Jr. (1980). *Applied time series analysis for the social sciences*. Beverly Hills, CA: Sage.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156-166.
- Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy*, *34*, 189-211.
- Parker, R. I., & Brossart, D. F. (2006). Phase contrasts for multiphase single case intervention designs. *School Psychology Quarterly*, *21*, 46-61.
- Parker, R. I., Cryer, J., & Byrns, G. (2006). Controlling baseline trend in single-case research. *School Psychology Quarterly*, *21*, 418-443.
- Parker, R. I., & Hagan-Burke, S. (2007). Median-based overlap analysis for single case data: A second study. *Behavior Modification*, *31*, 919-936.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. J. (2007). Percentage of all non-overlapping data: An alternative to PND. *Journal of Special Education*, *40*, 194-204.
- Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*, 357-367.
- Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, *75*, 135-150.
- Perdices, M., & Tate, R. L. (2010). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychological Rehabilitation*, *19*, 904-927.

- Reichardt, C. S. (2006). The principle of parallelism in the design of studies to estimate treatment effects. *Psychological Methods, 11*, 1-18.
- Robey, R. R., Schultz, M. C., Crawford, A. B., & Sinner, C. A. (1999). Single-subject clinical outcome research: designs, data, effect sizes, and analysis. *Aphasiology, 13*, 445-473.
- Schneider, N., Goldstein, H., & Parker, R. I. (2008). Social skills interventions for children with autism: A meta-analytic application of percentage of all non-overlapping data (PAND). *Evidence-Based Communication Assessment and Intervention, 2*, 152-162.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8*, 24-33.
- Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention, 2*, 188-196.
- Solanas, A., Manolov, R., & Onghena, P. (2010). Estimating slope and level change in N=1 designs. *Behavior Modification, 34*, 195-218.
- White, D. M., Rusch, F. R., Kazdin, A. E., & Hartmann, D. P. (1989). Applications of meta-analysis in individual subject research. *Behavioral Assessment, 11*, 281-296.
- Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *Journal of Special Education, 44*, 18-29.

Appendix 1

Calculation of the four procedures tested: word explanations and application to fictitious data with three phase A scores (5, 6, and 8) and three phase B scores (11, 8, and 11). Note that the example and the computation comments refer to interventions aiming *increase* in the behavior of interest. Formulae and further information can be found in the original papers introducing the indices.

Improvement rate difference (Parker et al., 2009)

The procedure compares two proportions – the improvement rate in phase A (IR_A) and improvement rate in phase B (IR_B). This index is closely related to the Percentage of All Nonoverlapping Data (Parker et al., 2007) and is based on the identification of the smallest number of data points that have to be removed from either phase in order to eliminate all overlap. It is important to consider whether a data point is removed from the baseline, and counted as improved (increasing IR_A), or from the treatment phase, and counted as not improved (reducing IR_B) in order to construct the corresponding 2x2 table, which is basic for the computation. The calculation procedure described in Parker et al. (2009) is applied to the example data (phase A: 5, 6, 8; phase B: 11, 8, 11). One alternative for achieving complete nonoverlap is to remove the third score from the baseline. Thus, there would be one improved in phase A and IR_A would be equal to $1/3 \approx .33$. Since no data are removed from phase B, there are three improvements in this phase and $IR_B = 3/3 = 1.0$. The IRD is $1 -$

.33 = .67. An alternative way for eliminating overlap is to remove the second score from the treatment phase, so that there are no improvements in phase A and two improvements in phase B. In this way, the computation is $IRD = IR_B - IR_A = 2/3 - 0/3 \approx .67$, the same value as obtained previously. The summary of the two alternative ways of computing IRD is provided in Table 1.

Table 1. Results of the analysis of the data regarding data points needing removal to eliminate overlap between phases.

Data points	Computational alternative 1		Computational alternative 2	
	Phase A	Phase B	Phase A	Phase B
Improved	1	3	0	2
Not improved	2	0	3	1
Total	3	3	3	3

An IRD value of 1.00 reflects lack of overlap, whereas .50 indicates that half of the scores are overlapping and is interpreted as chance level improvement. In the current example the amount of overlap, and the treatment effect, can be labeled arbitrarily as “moderate”.

Nonoverlap of all pairs (Parker & Vannest, 2009)

The procedure compares each baseline score with each treatment phase score and, therefore, $n_A \cdot n_B$ comparisons are made. A nonoverlapping pair is one in which the treatment score is greater than the baseline score. This quantity is

divided by $n_A \cdot n_B$ and the obtained value is interpreted as “the probability that a score drawn at random from a treatment phase will exceed that of a score drawn at random from a baseline phase” (Parker & Vannest, 2009, p.359). Using the example data, only the pair formed by the third baseline score and the second treatment phase score does overlap. Thus, only 1 of the $3 \cdot 3$ pairs overlaps and since ties count as half of an overlap - $NAP = 8.5/9 \approx .94$. This result may indicate that it is quite probable that a randomly selected phase B data point would be greater than a randomly selected phase A data point.

Percentage of nonoverlapping corrected data (Manolov & Solanas, 2009)

The procedure applies PND (Scruggs et al., 1987) after introducing a data correction step intended to remove baseline linear trend from the whole data series. Firstly, the phase A data are differenced. In the example baseline, this operation leads to $6 - 5 = 1$ and $8 - 6 = 2$. The mean of the differenced scores, which are now $n_A - 1$ instead of n_A , is computed (1.5 in the example) and it represents an estimate of the linear trend present in the baseline. Afterwards, trend is removed from both baseline and treatment data subtracting the trend estimate multiplied by the dummy variable representing time (ranging from 1 to $n_A + n_B$) from each original score. Therefore, the original number of measurements in each phase is maintained. In the example, the trend estimate 1.5 is multiplied by the measurement time (1) and subtracted from the first score (1), thus, $5 - 1.5 \cdot 1 = 3.5$. For the second score, $6 - 1.5 \cdot 2 = 3$, and so on till the last score where $11 - 1.5 \cdot 6 = 2$. Finally, PND is applied to the

differenced data. In the differenced fictitious data (3.5, 3, and 3.5 in phase A and 5, .5, and 2 in phase B), only the first treatment phase differenced score is greater than all baseline differenced scores and, thus, PNCD \approx 33% indicating a somewhat smaller effect than the previous indices.

Slope and level change (Solanas et al., 2010)

The procedure removes baseline linear trend from the whole data series prior to estimating slope and level change separately and in that order. In that sense, the procedure can be considered parametric taking into account that it is designed to estimate the β_1 , β_2 , and β_3 parameters which are assumed to be in the basis of the data generation process (Huitema & McKean, 2000), although no statistical decision is made (e.g., testing the null hypothesis that $\beta_2 = 0$). The only assumption made is that trend, if present, is approximately linear. Therefore, the application of the procedures is recommended only in cases when baseline data show relative stability or the data pattern resembles visually a linear upward or downward trend.

The first trend elimination step is identical to the one in PNCD and thus values used from now on are 3.5, 3, and 3.5 for phase A and 5, .5, and 2 for phase B. In a second step, the treatment phase data are differenced and their average is computed. Differencing in the example leads to $.5 - 5 = -4.5$ and $2 - 0.5 = 1.5$. The average quantifies the phase B trend and represents the change in slope, given that the baseline is assumed to have a zero slope after the first step. In this case the average is -1.5 indicating a negative change in

slope, i.e., the slope becomes less steep in the treatment phase by 1.5 measurement units at each observation. Afterwards, the phase B trend is removed from treatment data subtracting the trend estimate multiplied by a dummy variable ranging from 0 to $n_B - 1$ from each original score. Removing trend from the phase B in the example leads to $5 - (-1.5) \cdot 0 = 5$, $.5 - (-1.5) \cdot 1 = 2$, and $2 - (-1.5) \cdot 2 = 5$. After this manipulation phase B is also assumed to have zero trend and the means of the two phases are compared. Phase A data from which baseline trend was removed in the initial step are 3.5, 3, and 3.5 and their mean is approximately 3.33. Phase B data from each both baseline and treatment phase trends are removed are 5, 2, and 5, and their mean is 4. Therefore, the level change is $4 - 3.33 = .67$. This value suggests that, after controlling slope change, there is a slight positive level change – in average, the scores in the treatment phase are higher with less than a measurement unit (i.e., the unit used to quantify the behavior).

Appendix 2

R code for correcting linear trend prior to applying the Nonoverlap of All Pairs index. The input required from the user is a “info.dat” archive with the vector of measurements, read using the expression `info <- array(scan(“info.dat”))` and the phase A length specified after `n_a <-`. The remaining code is pasted in the R console as it appears to obtain NAP-s value.

```
-----  
# Data input  
info <- array(scan("info.dat"))  
n_a <- 3  
  
#Function  
napcorr <- function(info,n_a) {  
  slength <- length(info)  
  n_b <- slength-n_a  
  phaseA <- info[1:n_a]  
  phaseB <- info[(n_a+1):slength]  
  # Estimate phase A trend  
  phaseAdiff <- c(1:(n_a-1))  
  for (iter in 1:(n_a-1))  
    phaseAdiff[iter] <- phaseA[iter+1] - phaseA[iter]  
  trendA <- mean(phaseAdiff)  
  # Remove phase A trend from the whole data series  
  phaseAdet <- c(1:n_a)  
  for (timeA in 1:n_a)  
    phaseAdet[timeA] <- phaseA[timeA] - trendA * timeA  
  phaseBdet <- c(1:n_b)  
  for (timeB in 1:n_b)  
    phaseBdet[timeB] <- phaseB[timeB] - trendA * (timeB+n_a)  
  # Compute overlaps  
  overlap <- 0  
  for (iter1 in 1:n_a)  
    for (iter2 in 1:n_b)  
      {if (phaseAdet[iter1] > phaseBdet[iter2]) overlap <- overlap + 1;  
       if (phaseAdet[iter1] == phaseBdet[iter2]) overlap <- overlap + 0.5}  
  # Compute and print corrected NAP  
  nap <- ((n_a*n_b - overlap) / (n_a*n_b))*100  
  print ("NAP = "); print(nap)  
  overlap <- 0  
  list(nap)}  
  
# Magnitude of effect estimates  
results <- napcorr(info,n_a)  
-----
```

Figures

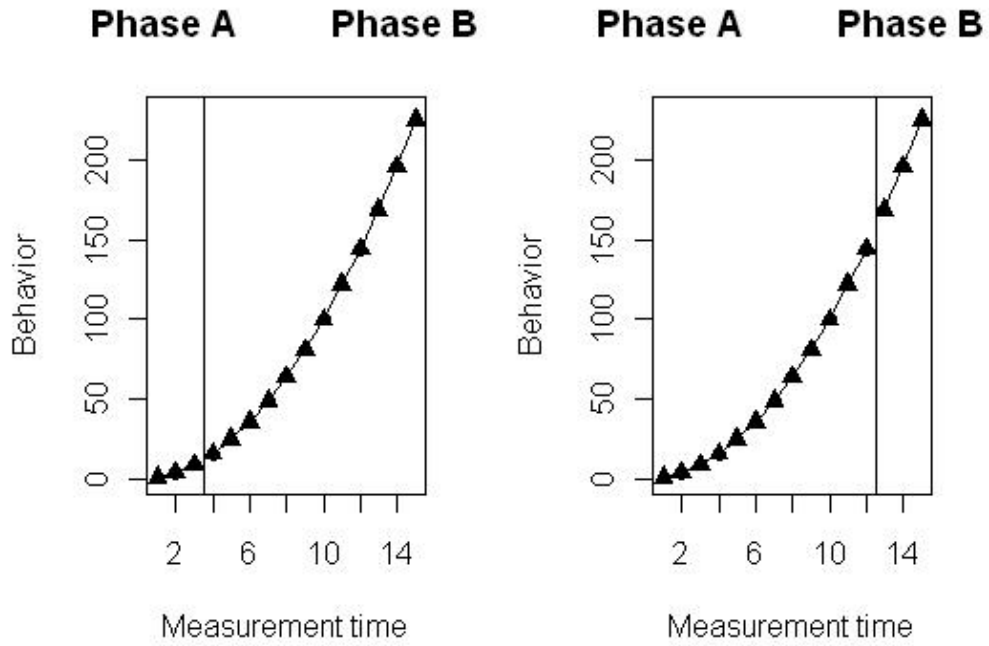


Figure 1. Ideal quadratic trend in an AB design, with $n_A = 3$ and $n_B = 12$ in the left panel and $n_A = 12$ and $n_B = 3$ in the right panel.

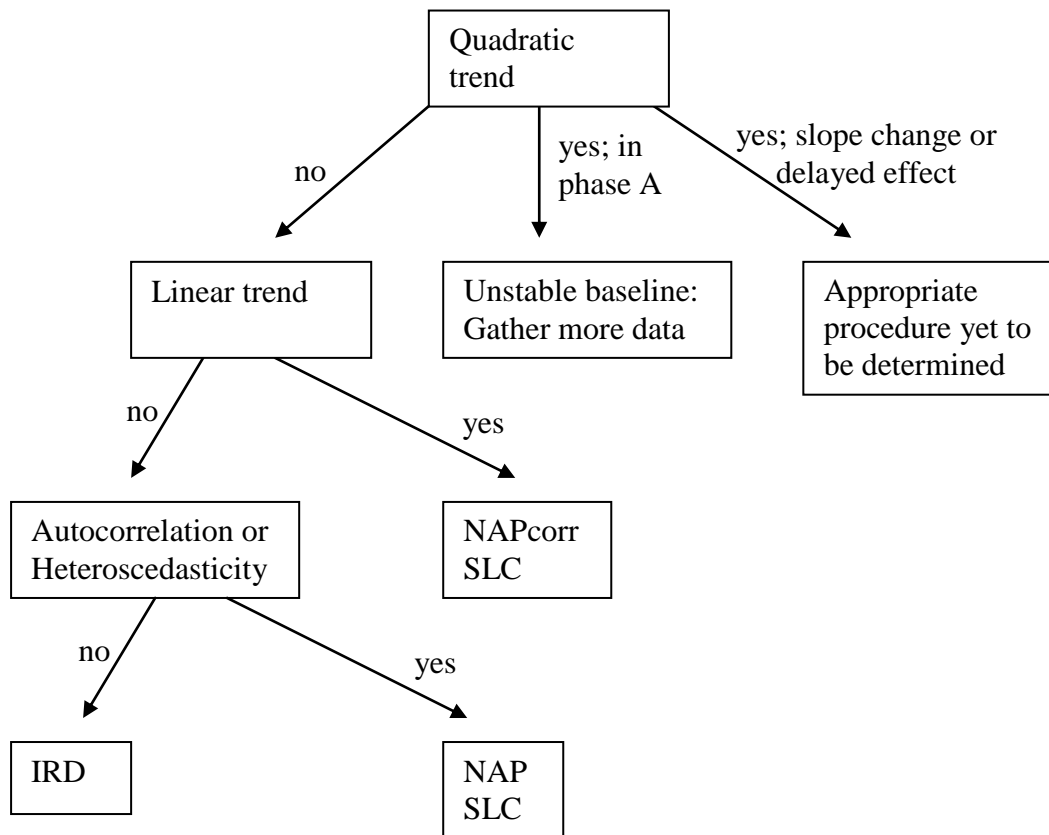


Figure 2. Flowchart for choosing among techniques according to the data characteristics identified by visual inspection. IRD denotes Improvement Rate Difference; NAP denotes Nonoverlap of All Pairs; NAPcorr denotes Nonoverlap of All Pairs with initial data correction step; and SLC denotes Slope and Level Change.

Tables

Table 1. Distortion due to autocorrelation for an AR(1) process – the values represent the ratio between serially dependent data ($\varphi_1 \neq 0$) and independent series ($\varphi_1 = 0$) in experimental conditions with no trend ($\beta_1 = 0$) or intervention effect ($\beta_2 = 0$ and $\beta_3 = 0$). The two values of Slope and Level Change estimators (not presented here) are equal to zero, regardless of the value of φ_1 .

Phase length		Ratio $\varphi_1 = -.3 / \varphi_1 = 0$			Ratio $\varphi_1 = .3 / \varphi_1 = 0$			Ratio $\varphi_1 = .6 / \varphi_1 = 0$		
n_A	n_B	IRD ¹	NAP ²	PNCD ³	IRD	NAP	PNCD	IRD	NAP	PNCD
exponential										
5	5	1.00	1.01	.94	1.07	.99	1.12	1.14	.96	1.25
10	10	1.01	1.00	.94	1.12	1.00	1.16	1.39	.99	1.48
15	15	1.03	1.00	.96	1.14	1.00	1.16	1.50	.99	1.53
normal										
5	5	.97	1.00	.95	1.08	1.00	1.06	1.19	1.00	1.12
10	10	.97	1.00	.94	1.13	1.00	1.06	1.40	1.00	1.20
15	15	.98	1.00	.96	1.13	1.00	1.07	1.47	1.00	1.22
uniform										
5	5	.96	1.00	.95	1.05	1.00	1.05	1.14	1.00	1.10
10	10	.96	1.00	.94	1.11	1.00	1.05	1.34	1.00	1.14
15	15	.98	1.00	.92	1.14	1.00	1.02	1.44	1.00	1.09

¹ Improvement Rate Difference.

² Nonoverlap of All Pairs.

³ Percentage of Nonoverlapping Corrected Data.

Table 2. Distortion due to autocorrelation for an MA(1) process – the values represent the ratio between serially dependent data ($\theta_1 \neq 0$) and independent series ($\theta_1 = 0$) in experimental conditions with no trend ($\beta_1 = 0$) or intervention effect ($\beta_2 = 0$ and $\beta_3 = 0$). The two values of Slope and Level Change estimators (not presented here) are equal to zero, regardless of the value of θ_1 .

Phase length		Ratio $\theta_1=.3 / \theta_1=0$			Ratio $\theta_1=-.3 / \theta_1=0$			Ratio $\theta_1=-.9 / \theta_1=0$		
n_A	n_B	(Ratio $\phi_1 \approx -.27 / \phi_1=0$)			(Ratio $\phi_1 \approx .27 / \phi_1=0$)			(Ratio $\phi_1 \approx .49 / \phi_1=0$)		
exponential		IRD ¹	NAP ²	PNCD ³	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	.99	1.01	.91	1.06	.99	1.11	1.16	.98	1.25
10	10	.99	1.00	.93	1.08	1.00	1.12	1.27	.99	1.31
15	15	1.01	1.00	.92	1.09	1.00	1.11	1.29	.99	1.30
normal		IRD	NAP	PNCD	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	.96	1.00	.96	1.07	1.00	1.05	1.17	1.00	1.12
10	10	.96	1.00	.94	1.09	1.00	1.05	1.24	1.00	1.08
15	15	.97	1.00	.94	1.10	1.00	1.05	1.26	1.00	1.06
uniform		IRD	NAP	PNCD	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.05	1.00	.96	1.14	1.00	1.05	1.23	1.00	1.11
10	10	.94	1.00	.94	1.08	1.00	1.03	1.18	1.00	1.03
15	15	.95	1.00	.92	1.09	1.00	1.01	1.18	1.00	.95

¹ Improvement Rate Difference.

² Nonoverlap of All Pairs.

³ Percentage of Nonoverlapping Corrected Data.

Table 3. Distortion due to increased or decreased variability in phase B – the values represent the ratio between heteroscedasticity and homoscedasticity in independent data series ($\varphi_I = 0$ or $\theta_I = 0$) with no treatment effect ($\beta_2 = 0$ and $\beta_3 = 0$). The values of the two Slope and Level Change estimators (not presented here) are equal to zero, regardless of the variability in the series.

Phase length		Ratio increased variability /			Ratio decreased variability /		
n_A	n_B	homoscedastic data series			homoscedastic data series		
exponential		IRD ¹	NAP ²	PNCD ³	IRD	NAP	PNCD
5	5	.91	.81	1.01	1.80	1.19	1.00
10	10	1.12	.81	1.20	2.89	1.19	.88
15	15	1.34	.81	1.37	4.08	1.19	.82
normal		IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.21	1.00	1.12	1.22	1.00	.94
10	10	1.60	1.00	1.29	1.60	1.00	.86
15	15	2.00	1.00	1.89	2.00	1.00	1.07
uniform		IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.33	1.00	1.16	1.33	1.00	.93
10	10	2.02	1.00	1.35	2.02	1.00	.84
15	15	2.76	1.00	1.46	2.77	1.00	.79

¹ Improvement Rate Difference.

² Nonoverlap of All Pairs.

³ Percentage of Nonoverlapping Corrected Data.

Table 4. Distortion due to linear trend – the values represent the ratio between $\beta_1 \neq 0$ and $\beta_1 = 0$ in independent data series ($\varphi_1 = 0$ or $\theta_1 = 0$) with no effect ($\beta_2 = 0$ and $\beta_3 = 0$). The values of the two Slope and Level Change estimators (not presented here) are equal to zero, regardless of the value of β_1 .

Phase length		Ratio $\beta_1=.1 / \beta_1=0$			Ratio $\beta_1=.2 / \beta_1=0$			Ratio $\beta_1=.3 / \beta_1=0$		
n_A	n_B	IRD ¹	NAP ²	PNCD ³	IRD	NAP	PNCD	IRD	NAP	PNCD
exponential		IRD ¹	NAP ²	PNCD ³	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.98	1.38	1.01	2.57	1.60	1.00	2.95	1.73	1.00
10	10	3.96	1.60	1.00	5.20	1.81	.99	5.79	1.90	1.00
15	15	6.46	1.73	1.01	8.17	1.90	1.00	8.86	1.95	1.02
normal		IRD	NAP	PNCD	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.58	1.27	1.01	2.16	1.50	1.00	2.65	1.67	1.00
10	10	2.86	1.50	.99	4.62	1.78	1.00	5.55	1.89	1.00
15	15	4.75	1.67	1.01	7.53	1.89	1.00	8.54	1.95	1.00
uniform		IRD	NAP	PNCD	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.60	1.26	.99	2.11	1.48	.99	2.55	1.64	.99
10	10	3.00	1.48	1.00	4.50	1.77	1.00	5.46	1.89	1.00
15	15	4.92	1.65	1.00	7.49	1.89	1.00	8.51	1.95	1.00

¹ Improvement Rate Difference.

² Nonoverlap of All Pairs.

³ Percentage of Nonoverlapping Corrected Data.

Table 5. Distortion due to quadratic trend – the values represent the ratio between $\beta_1 \neq 0$ and $\beta_1 = 0$ in independent data series ($\varphi_1 = 0$ or $\theta_1 = 0$) with no effect ($\beta_2 = 0$ and $\beta_3 = 0$).

Phase length		Ratio $\beta_1=.01 / \beta_1=0$				
n_A	n_B					
exponential		IRD ¹	NAP ²	PNCD ³	SLC-LC ⁴	SLC-SC ⁴
5	5	1.99	1.40	1.14	.04	.10
10	10	5.05	1.82	1.91	.10	.20
15	15	8.59	1.95	3.66	.16	.30
normal		IRD	NAP	PNCD	SLC-LC	SLC-SC
5	5	1.64	1.30	1.13	.05	.10
10	10	4.68	1.78	1.80	.10	.20
15	15	8.59	1.95	3.23	.16	.30
uniform		IRD	NAP	PNCD	SLC-LC	SLC-SC
5	5	1.66	1.29	1.11	.05	.10
10	10	4.64	1.77	1.77	.10	.20
15	15	8.55	1.95	2.98	.14	.30

¹ Improvement Rate Difference.

² Nonoverlap of All Pairs.

³ Percentage of Nonoverlapping Corrected Data.

⁴ Level change (LC) and slope change (SC) estimators of the Slope and Level Change. The values for the distortion are expressed as absolute differences from the values for $\beta_1=0$, since the latter are equal to zero and a ratio cannot be computed.

Table 6. Detection of level change – the values represent the ratio between $\beta_2 \neq 0$ and $\beta_2 = 0$ in independent data series ($\varphi_1 = 0$ or $\theta_1 = 0$) with no trend ($\beta_1 = 0$) or slope change ($\beta_3 = 0$). The average value yielded by the level change estimator of Slope and Level Change is equal to β_2 (.1, .2, and .3) and the slope change estimator average is equal to $\beta_3 = 0$.

Phase length		Ratio $\beta_2=.1 / \beta_2=0$			Ratio $\beta_2=.2 / \beta_2=0$			Ratio $\beta_2=.3 / \beta_2=0$		
n_A	n_B									
exponential		IRD ¹	NAP ²	PNCD ³	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.24	1.10	1.06	1.46	1.18	1.13	1.67	1.26	1.20
10	10	1.45	1.10	1.08	1.90	1.18	1.16	2.33	1.26	1.23
15	15	1.69	1.10	1.08	2.40	1.18	1.17	3.08	1.26	1.25
normal		IRD	NAP	PNCD	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.11	1.06	1.06	1.22	1.11	1.11	1.35	1.17	1.16
10	10	1.14	1.06	1.07	1.30	1.11	1.15	1.48	1.17	1.23
15	15	1.16	1.06	1.09	1.35	1.11	1.18	1.56	1.17	1.27
uniform		IRD	NAP	PNCD	IRD	NAP	PNCD	IRD	NAP	PNCD
5	5	1.13	1.06	1.05	1.26	1.11	1.10	1.39	1.17	1.15
10	10	1.24	1.05	1.07	1.48	1.11	1.15	1.70	1.17	1.23
15	15	1.35	1.06	1.09	1.68	1.11	1.18	2.01	1.16	1.26

¹ Improvement Rate Difference.

² Nonoverlap of All Pairs.

³ Percentage of Nonoverlapping Corrected Data.

Table 7. Detection of slope change – the values represent the ratio between $\beta_3 \neq 0$ and $\beta_3 = 0$ in independent data series ($\varphi_1 = 0$ or $\theta_1 = 0$) with no trend ($\beta_1 = 0$) or level change ($\beta_2 = 0$). The average value yielded by the slope change estimator of Slope and Level Change is equal to β_3 (.1, .2, and .3) and the level change estimator average is equal to $\beta_2 = 0$.

Phase length		Ratio $\beta_3=.1 / \beta_3=0$			Ratio $\beta_3=.2 / \beta_3=0$			Ratio $\beta_3=.3 / \beta_3=0$		
n_A	n_B	IRD ¹	NAP ²	PNCD ³	IRD	NAP	PNCD	IRD	NAP	PNCD
exponential										
5	5	1.39	1.18	1.12	1.66	1.30	1.23	1.88	1.40	1.35
10	10	2.19	1.34	1.37	2.87	1.52	1.79	3.45	1.63	2.25
15	15	3.09	1.46	1.69	4.35	1.65	2.60	5.44	1.75	3.53
normal										
5	5	1.23	1.11	1.10	1.45	1.22	1.20	1.68	1.31	1.30
10	10	1.74	1.25	1.33	2.60	1.44	1.70	3.40	1.58	2.06
15	15	2.48	1.36	1.66	4.35	1.60	2.40	5.80	1.72	3.08
uniform										
5	5	1.25	1.11	1.08	1.48	1.21	1.18	1.68	1.30	1.29
10	10	1.95	1.24	1.33	2.74	1.42	1.68	3.49	1.57	2.03
15	15	2.99	1.35	1.62	4.75	1.58	2.28	6.23	1.71	2.86

¹ Improvement Rate Difference.

² Nonoverlap of All Pairs.

³ Percentage of Nonoverlapping Corrected Data.

Table 8. Summary of effect size quantification for $\beta_2 = \beta_3 = 0$. “ND” denotes no distortion in the quantification; a desirable feature. “U” and “O” denote underestimation and overestimation, respectively, being both considered as drawbacks.

Procedure tested	Autocorrelation		Phase B variability		Trend	
	negative	positive	increased	reduced	linear	quadratic
IRD ¹	U	O	O	O	O	O
NAP ²	ND	ND	ND	ND	O	O
PNCD ³	U	O	O	U	ND	O
SLC ⁴	ND	ND	ND	ND	ND	O

¹ Improvement Rate Difference.

² Nonoverlap of All Pairs.

³ Percentage of Nonoverlapping Corrected Data.

⁴ Slope and Level Change.