

## **Linear trend in single-case visual and quantitative analyses**

Rumen Manolov

Affiliation: Department of Social Psychology and Quantitative Psychology, Faculty of Psychology, University of Barcelona, Spain

Correspondence: Rumen Manolov, Departament de Psicologia Social i Psicologia Quantitativa (Secció de Psicologia Quantitativa), Facultat de Psicologia, Universitat de Barcelona, Passeig de la Vall d'Hebron, 171, 08035-Barcelona, Spain. Phone number: +34934031137. E-mail: [rrumenov13@ub.edu](mailto:rrumenov13@ub.edu)

Acknowledgements: The author would like to thank Dr. Patrick Onghena for his insightful feedback on previous versions of this text.

**Abstract**

The frequently used visual analysis of single-case data focuses on data aspects such as level, trend, variability, overlap, immediacy of effect, and consistency of data patterns; most of these aspects are also commonly quantified besides inspecting them visually. The present text focuses on trend, because even linear trend can be operatively defined in several different ways, while there also different approaches for controlling for baseline trend. We recommend using a quantitative criterion for choosing a trend line fitting technique and comparing baseline and intervention slopes, instead of detrending. We implement our proposal it in a free web-based application created specifically for following the What Works Clearinghouse *Standards* recommendations for visual analysis. This application is especially destined to applied researchers and provides graphical representation of the data, visual aids, and quantifications of the difference between phases in terms of level, trend, and overlap, as well as two quantifications of the immediate effect. An evaluation of the consistency of effects across replications of the AB sequence is also provided. For methodologists and statistician, we include formulas and examples of the different straight line fitting and detrending techniques in order to improve the reproducibility of results and simulations.

*Keywords:* single-case designs, trend, visual analysis, software

Running head: SINGLE-CASE LINEAR TREND

Single-case designs (SCD) data are still commonly analyzed visually (Campbell & Herzinger, 2010; Kratochwill, Levin, Horner, & Swoboda, 2014; Parker, Cryer, & Byrns, 2006; Smith, 2012), either on exclusive basis or jointly with an objective quantification. When describing SCD visual analysis, the What Works Clearinghouse (WWC) technical report (Kratochwill et al., 2010) mentions six aspects of the data: level, trend, overlap, variability, immediacy of effect, and consistency of data patterns across similar conditions. These data aspects can be assessed purely visually, but they are also frequently quantified. Specifically, level is usually represented by the mean and the median (Fisher, Kelley, & Lomas, 2003; Lane & Gast, 2014); trend is defined as the “best fitting straight line” (Kratochwill et al., 2010); variability can be represented using standard deviation bands (Fisher et al., 2003; Pfadt & Wheeler, 1995), range and the stability envelope (Lane & Gast, 2014); and overlap is commonly assessed in visual analysis via the Percentage of nonoverlapping data (PND; Lane & Gast, 2014). PND is also frequently used as main quantification of effect (Schlosser, Lee, & Wendt, 2008; Scruggs & Mastropieri, 2013), despite its limitations (Allison & Gorman, 1994; Campbell, 2013), and despite the number of nonoverlap alternatives that have already been discussed extensively elsewhere (Parker, Vannest, & Davis, 2011; Wolery, Busick, Reichow, & Barton, 2010). The immediacy of the effect can be assessed focusing on the level of last three baseline measurements versus the first three intervention phase measurements (Kratochwill et al., 2010). Regarding the final data aspect, the WWC *Standards* define the consistency of similar phases as the “extent to which the data patterns in phases with the same (or similar) procedures are similar” (Kratochwill et al., 2010, p. 20). We refer to this topic in a later section entitled “Consistency of Effects and Type of Single-Case Design”.

### **The Importance of Trend**

The focus of the current text is on trend, due the evidence of the importance of trend lines for improving the performance of visual analysts (Bailey, 1984; Skiba, Deno, Marston, & Casey, 1989) and given that neither the *WWC Standards*, nor Horner Swaminathan, Sugai, and Smolkowski (2012) mention a specific technique for fitting trend. Trend has been the focus of attention in relation to both visual analysis (Mercer & Sterling, 2012) and statistical analytical techniques (e.g. Parker et al., 2006; Solomon, 2014), but even “linear trend” is an ambiguous term, unless clearly defined. Actually, to the best of our knowledge, there has been no broad review and no systematic comparison performed on the different ways in which a straight line can be fitted to the single-case data. In order to underline the importance of trend, we can relate it to the other data aspects mentioned in the *WWC Standards*. First, a trend line is a less restrictive visual representation of the data than a median line, which imposes lack of change in level with time, or stationarity. Second, the trend line can be used as a basis of the visual representation and the quantification of variability (i.e., if variability is defined as the amount of variation around the trend line). Third, as will be reviewed in depth in Appendix A, some nonoverlap indices take trend into account, in order to provide a more meaningful comparison between conditions. Finally, the immediate effect of the intervention can also be conceptualized and quantified on the basis of the projection from the baseline (Horner et al., 2012), for instance, as the difference between the predictions for the first intervention phase value from the baseline and the intervention phase trend lines, as in piecewise regression (Center, Skiba, & Casey, 1985-1986). Thus, we consider that trend is crucial data aspect that requires thorough discussion.

### **Consistency of Effects and Type of Single-Case Design**

The consistency of effects can be assessed in relation to the type of single-case design used. Specifically, although comparing level, trend, and variability and assessing overlap can be done separately for each AB-pair of phases, an AB design is not sufficiently rigorous from a methodological perspective. Actually, the *WWC Standards* (Kratochwill et al., 2010) require a minimum of three replications of the effect that can be observed in an AB-pair. However, not all replications are equivalent. For instance, focusing on two recently compared designs (Novotny et al., 2014), a nonconcurrent multiple-baseline design would entail three AB comparisons, whereas an ABAB design entails comparing  $A_1$ - $B_1$ ,  $B_1$ - $A_2$ , and  $A_2$ - $B_2$ . Therefore, what is different is that in an ABAB design: (a) there is a single participant and potentially more consistency can be expected; (b) one of the comparisons is in the BA order, which might affect the visual impression of the results; (c) the information from two of the phases is used twice, which could lead to apparently greater consistency; and (d) in some cases,  $A_2$  may not be considered comparable to  $A_1$ , because the effect of the intervention may not wash out completely, which could lead to lower consistency. It is even more relevant to distinguish between a concurrent and a nonconcurrent multiple-baseline design across participants (Carr, 2005), given that for the former the staggered introduction of the interventions is crucial for assessing experimental control via within-series and between-series comparisons (Ferron, Moeyaert, Van Den Noortgate, & Beretvas, 2014; Horner et al., 2005). Therefore, even if there is similar consistency in AB-comparisons arising from these two designs, the evidence from a concurrent multiple-baseline design is stronger. In summary, it has to be underlined that in absence of replication and consistency of the effect, a difference observed in an AB-comparison (in terms of level,

Running head: SINGLE-CASE LINEAR TREND

trend, variability or overlap) is insufficient. Moreover, the assessment of consistency across AB-comparisons has to take place considering the type of SCD actually used.

### **Aims of the Article, Justification, Intended Audience and Organization**

The article aims to provide an answer to several questions. As presented in Table 1, we consider that some of these questions (How to choose a technique for fitting a straight line?, Is it necessary to always fit a (straight) line?, Is it necessary to always control for baseline trend?, How to follow the WWC *Standards* recommendations for visual analysis?) might be of greater interest to applied researchers. The ultimate aim is that applied researchers have a tool enabling them to perform visual analysis, focusing on each of the six data aspects mentioned in the WWC *Standards* in a systematic way. For that reason, we have included our comments and recommendations on these questions in the main text.

**Table 1.** Questions aimed to be answered, intended audience, and organization of the text.

<b>Question</b>	<b>Intended audience</b>	<b>Content availability</b>	<b>Approach to answering the question</b>	<b>Summary for applied researchers</b>
What trend line fitting techniques exist and how do they differ?	Methodologists / statisticians	Main text (verbal explanations) and Appendix A (formulas and examples)	Review of trend line fitting techniques incorporated in single-case data analysis procedures	There are several options that do not necessarily lead to the same result
How to choose a technique for fitting a straight line?	Applied researchers; Methodologists / statisticians	Main text (verbal explanations) and Appendix B (formulas and examples)	Search in the literature for a quantification of fit of a trend line to the data	Use the Mean Absolute Scaled Error criterion
Is it necessary to always fit a (straight) line?	Applied researchers	Main text	Review of alternatives and a recommendation	No, but it is easier, more parsimonious, more common
What ways of controlling for baseline trend have been incorporated in single-case data analytical procedures?	Methodologists / statisticians	Main text (verbal explanations) and Appendix C (formulas and examples)	Review of existing approaches to controlling for baseline trend	There are several options that do not necessarily lead to the same result
Is it necessary to always control for baseline trend?	Applied researchers	Main text	Discussion and a recommendation	Baseline trend control is not the same as detrending; separate trend lines can be fit
How to follow the WWC <i>Standards</i> recommendations for visual analysis?	Applied researchers	Main text	Creation of a web application	Use the web application created

Running head: SINGLE-CASE LINEAR TREND

We also aim to provide answers to questions we might be of greater interest to methodologists and statisticians (What trend line fitting techniques exist and how do they differ?, What ways of controlling for baseline trend have been incorporated in single-case data analytical procedures?). Thus, we provide formulas and examples on the existing options for fitting trend lines and for detrending in Appendices A and B. We consider that a single document containing this information is relevant and necessary, because otherwise this information is either missing (for Tau-U and Baseline corrected Tau) or scattered in several documents, including difficult to access sources. For applied researchers not interested in the details, the brief verbal comments in the main text and the answers provided in the last column of Table 1 would be sufficient.

### **What Trend Line Fitting Techniques Exist and How Do They Differ?**

The current overview is focused on techniques that have been incorporated in analytical procedures proposed for single-case data analysis. Thus, several trend estimation techniques are not included, such as the minimum m-estimation (Anderson & Schumacker, 2003; Yohai, 1987) tested by Brossart, Parker, and Castillo (2011) but not included in a data analytical procedure, or the least median of squares (see Wilcox, 2012, Chapter 10). The current review is also focused on linear trend here rather than on nonlinear trend (commented more extensively in the section entitled “Is It Necessary to Always Fit A (Straight) Line?”), because it is simpler, potentially more easily understood (Chatfield, 2000), and also more frequently present in visual aids (Fisher et al., 2003; Lane & Gast, 2014). Thus, we prioritize the accessibility of the techniques (Parker, Vannest, & Davis, 2014). Moreover, in the WWC *Standards* linear trends are



Running head: SINGLE-CASE LINEAR TREND

the focus of attention, according to the illustrations provided in Kratochwill et al. (2010).

Apart from stating that there are many more trend line fitting techniques than the *linear* trend line fitting techniques incorporated in single-case data analytical procedures<sup>1</sup>, the answer to first part of the question is provided in Table 2. As seen in Table 2, least squares estimation is the technique most frequently incorporated in single-case data analytical procedures. Moreover, some analytical procedures model only baseline trend (for detrending or for comparing with the intervention phase data), whereas the majority of procedures using least squares estimation fit separate trend lines to the baseline and the intervention phases.

---

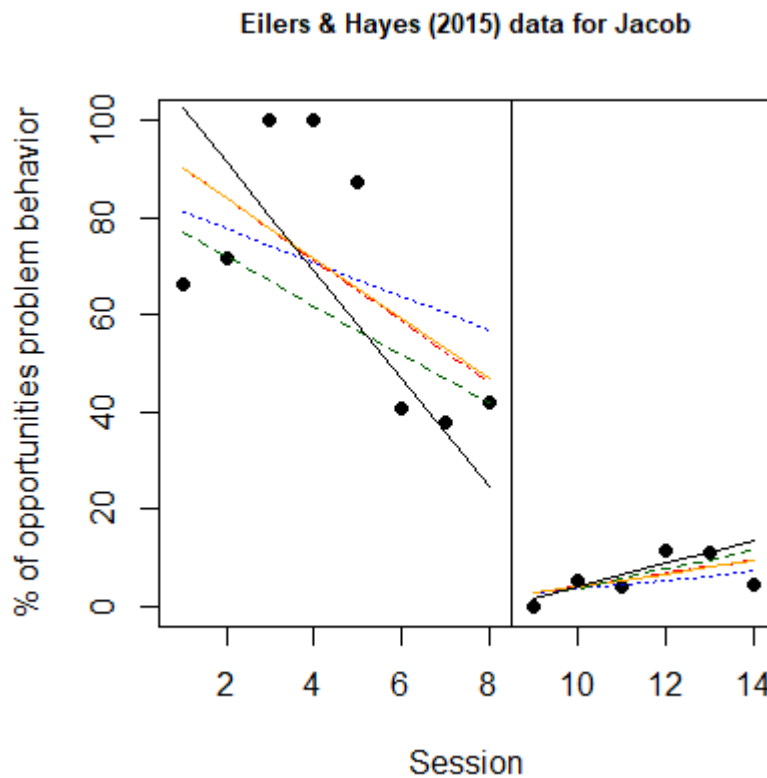
<sup>1</sup> It can be verified that the analytical procedures mentioned cover the procedures listed in previous broad overviews (e.g., Gage & Lewis, 2013; Manolov & Moeyaert, 2017; see also Tate et al.'s, 2016, Appendix), considering that the focus here is only on procedures incorporating trend estimation.

Running head: SINGLE-CASE LINEAR TREND

**Table 2.** Cross tabulation of straight line fitting techniques (columns) and the portion of the data to which the trend line is fitted (rows)

Where trend is modeled	Least squares estimation	Bi-split (split-middle method)	Tri-split	Theil-Sen estimator	Differencing
Baseline only	Allison and Gorman (1993) ordinary least squares regression model	Conservative Dual Criterion (Fisher et al., 2003); Percentage of data points exceeding median trend (Wolery et al., 2010)	Graph rotation procedure (Parker et al., 2014)	Baseline corrected Tau (Tarlow, 2017)	Mean Phase Difference (Manolov & Rochat, 2015; Manolov & Solanas, 2013)
Baseline and intervention phase separately	Last Treatment Day procedure (D. M. White, Rusch, Kazdin, & Hartmann, 1989); Piecewise regression (Center et al., 1985-1986); Mean and Slope Adjustment (MASAJ) mean plus trend shift (Parker et al., 2006); Generalized least squares regression proposal by Swaminathan et al. (2014); Standardized mean difference based on multilevel modeling of level and trend (Pustejovsky et al., 2014)	Visual analysis (Lane & Gast, 2014; Miller, 1985)	None	None	Slope and Level Change (Solanas et al., 2010)
Baseline and intervention phase jointly	Gorsuch's (1983) trend analysis	None	None	None	None

Regarding the second part of the question (“How do the trend line fitting techniques differ?”), from an applied perspective it is sufficient to say that greater differences between the trend lines fitted are expected when the data conform worse to a straight line (as is the case in the A phase in Figure 1). Complementarily, smaller differences are expected when the data are more easily approximated by a straight line (as is the case in the B phase in Figure 1).



**Figure 1.** Data (obtained by Eilers and Hayes, 2015) for Jacob (tier 3), a participant with autism spectrum disorder, treated with cognitive defusion exercise and exposure therapy. Black continuous line: bi-split technique for estimating trend; Light grey solid line: tri-split technique for estimating trend; dash-and-dot line: ordinary least squares; dotted line: differencing; dark grey dashed line: Theil-Sen. The values before the vertical line refer to the baseline (A) phase, whereas the values after the vertical line refer to the intervention (B) phase.

In brief, ordinary least squares estimation fits the trend line in such a way that the squared difference between itself and the actual values is the smallest possible. In comparison to the ordinary least squares estimation, tri-split and the Theil-Sen estimator are intended to be more resistant to (i.e., less affected by) outliers. In contrast to the previously mentioned techniques, the main advantage of the split-middle method is its simplicity and possibility to use it manually (Miller, 1985). Finally, differencing can be used for estimating trend as the average of the differences between successive data points, and not only for detrending, as it will be mentioned later. For further detail, a technical and formulaic representation of the five straight line fitting techniques is available in Appendix A.

### **How to Choose a Technique for Fitting a Straight Line?**

We consider that it is important to reach the same conclusion about intervention effect when focusing on the same data feature (here, trend). However, reaching the same conclusion is not ensured in light of the variety of straight line fitting techniques summarized in Table 2. Thus, it is not desirable that an apparently clear criterion such as “linear trend” would lead to very different graphical representations (superimposed lines) and numerical results. Accordingly, in visual analysis the need for formal decision rules has been underscored (Ottenbacher, 1990). Moreover, Parsonson and Baer (1992) explicitly mention the importance of both comparing trend line fitting techniques and having a rule for such a comparison.

A logical option for choosing between techniques is on the basis of the reliability (Parker, Vannest, Davis, & Sauber, 2011) or the amount of fit of the trend line to the data. It can logically be expected that, for the same amount of residual variability, more reliable estimates will be obtained for longer data series. The amount of fit of the

Running head: SINGLE-CASE LINEAR TREND

straight line to the data can be quantified in several ways, the most well-known of which is  $R^2$ , quantifying the proportion of data variability explained by a linear.

Actually, Hyndman and Koehler (2006) compare several options for comparing the actual values to the fitted values (e.g., the trend line) and they propose a measure called the Mean Absolute Scaled Error (MASE). Values of MASE greater than one indicate that the trend line provides a worse fit than predicting each value from the previous one.

In general, the greater the MASE, the worse the fit of the trend line to the data. In

Appendix B, the formulas for several measures of fit, including the MASE recommended here, are provided.

### **Is It Necessary to Always Fit A (Straight) Line?**

A straight line not always easily or meaningfully represents data. In general, it may not be always visually clear whether: (a) a straight line represents the data well; (b) a straight or a curved line is a better representation; and (c) which curved line to choose.

For simple monotonic relations in which there is only one bend in the curve representing the relation between the measurement occasions and the measurements, it is possible to use Mosteller and Tukey's bulging rule (see e.g. Fox, 2016, p. 66) for transforming the data and using afterwards any of the techniques for finding a best-fitting straight line. However, single-case researchers may not be willing to transform the data, as it goes against the importance of an in-depth understanding of the values obtained (Fahmie & Hanley, 2008). Without the need to transform, three options have been mentioned in the literature in order to model nonlinear trends: (a) polynomial regression (Swaminathan, Rogers, Horner, Sugai, & Smolkowski, 2014), considering that Horner et al. (2012) explicitly mention quadratic lines; (b) local regression (Solmi, Onghena, Salmaso, & Bulté, 2014); and (c) generalized additive models: applying cubic regression splines with many knots (e.g., 10 knots for a data series of 15 measurements;

Running head: SINGLE-CASE LINEAR TREND

Sullivan, Shadish, & Steiner, 2015). These two latter options are also useful when a nonlinear relation presents more than one bend, unlike quadratic models.

We consider that there are three criteria for selecting the appropriate model. First, in case there is substantive knowledge on the expected evolution of the target behavior (e.g., about the type of natural recovery during baseline, about the presence of an upper or a lower asymptote once the behavior has started improving after the intervention), such knowledge should be used. Second, scientific parsimony calls for using the simplest possible model that implies only a minor loss of information (or loss of goodness of fit) with respect to more complex models. In that sense, a linear model would be warranted in absence of a justification for a specific nonlinear model. Third, quantifications such as the  $R^2$  and MASE may be used for selecting a model of the data. Finally, it is also possible to use single-case analytical procedures that do not summarize the data via trend lines or mean lines, such as the Nonoverlap of all pairs (NAP; Parker & Vannest, 2009).

### **What Ways of Controlling for Baseline Trend Have Been Incorporated in Single-Case Data Analytical Procedures?**

Just as “linear trend” does not necessarily refer to a single possible straight line representing the relation between the measurement occasions and the measurements, “detrending” (i.e., removing baseline trend from the data) could also be an ambiguous term. Table 3 provides a brief summary of the different options for detrending, whereas in Appendix C we have included the formulas for each detrending technique and an example. Note that apart from using baseline trend for detrending, some analytical procedures extrapolate baseline trend for comparison purposes (e.g., Conservative dual

Running head: SINGLE-CASE LINEAR TREND

criterion; Percentage of data points exceeding median trend; Mean Phase Difference), whereas others compare the slopes of the trend lines fitted separately to the baseline and intervention phase (e.g., Piecewise and Generalized least squares regression, multilevel models).

**Table 3.** Approaches to detrending in several single-case data analytical procedures.

<b>Detrending approach</b>	<b>Single-case data analytical technique</b>	<b>Measurement units of the detrended data</b>
Initial baseline regression with time as predictor	Allison & Gorman (1993)	Transformed (residuals)
Initial regression with time and the interaction between time and phase as predictors	Before the Between-cases standardized mean difference (Shadish et al., 2014)	Transformed (residuals)
Single regression step modeling general trend and change in level	Trend analysis (Gorsuch, 1983)	Transformed (residuals)
Removing baseline trend from the whole series	Baseline corrected Tau Slope and Level Change	Original
Removing baseline trend from the intervention phase data	Mean and Slope Adjustment (Parker et al., 2006)	Original
Subtracting the number of improving baseline data points from the nonoverlap comparison	Tau-Y (Parker, Vannest, Davis, & Sauber, 2011)	Original
Differencing	Differencing analysis (Gorsuch, 1983) and ARIMA models (Harrop & Velicer, 1985)	Transformed

*Note.* ARIMA – autoregressive intergrated moving average.



### **Is It Necessary to Always Control for Baseline Trend?**

Parker et al. (2006) underscore the importance of taking baseline trend into account before quantifying intervention effects. However, it needs to be highlighted that taking baseline trend into account does not necessarily imply detrending (transforming the data), because baseline trend can be compared to the intervention phase trend. Actually, we here advocate for avoiding detrending by modeling separately the trend in each phase. In the next paragraph, we present the justification for our recommendation.

A first issue with detrending is that there are several different ways in which it can be achieved, as reviewed in the section entitled “What Ways of Controlling for Baseline Trend Have Been Incorporated in Single-Case Data Analytical Procedures?”. Thus, it would be necessary to have a criterion for choosing one of these options. A second issue is that it may not always be easy to decide for which data sets to detrend: always, only when baseline trend that is statistically significant, or only when the baseline trend represents well the data. A similar concern has been expressed by Parker, Vannest, Davis, and Sauber (2011), who underline that the length of the baseline, and the possibility of the data correction being excessively strong have to be considered. Moreover, detrending is not flawless. The evidence suggests that in some cases baseline trend control may be insufficient, leading to excessively large estimates of effect. For instance, Tarlow (2017) reports such evidence for Tau-U and also for Baseline corrected Tau, if the statistical significance of the baseline slope is assessed in baselines with fewer than 10 measurements (Tarlow, 2017). Complementarily, in other cases baseline trend control may be excessively strong control, leading to conservative estimates of effect (e.g., see Parker & Brossart, 2003, for Gorsuch’s, 1983, trend analysis and Manolov, Arnau, Solanas, & Bono, 2010, for differencing analysis). Similarly, in relation to the initial detrending before applying the BC-SMD, Shadish, Hedges, and

Running head: SINGLE-CASE LINEAR TREND

Pustejovsky (2014) underline that detrending should be used with caution, because it may remove part of the intervention effect when it is expressed as change in slope. More favorable results have been reported for the Allison and Gorman (1993) model (Manolov & Solanas, 2008; Tarlow, 2017), for the Mean Phase Difference (Tarlow, 2017), and for the Slope and Level Change procedure (Solanas et al., 2010). However, it is difficult to disentangle the performance of the detrending technique from the performance of the data analytical procedures in which it is incorporated.

Regarding our recommendation, the separate modelling of baseline and intervention phase trend is best aligned with the way in which visual analysis proceeds. We here suggest choosing the straight line fitting technique on the basis of the MASE. In terms of further quantifications, it is possible to compare trend lines by means of a randomization test (Michiels, Heyvaert, Meulders, & Onghena, 2017). Additionally, multilevel models also proceed estimating trend lines separately. For instance, an option to quantify the BC-SMD for data potentially presenting trend is to use multilevel models and maximum likelihood estimation (Pustejovsky, Hedges, & Shadish, 2014) instead of detrending first.

### **How to Follow the WWC *Standards* Recommendations for Visual Analysis?**

In the present section, we describe a web-based application available at <https://manolov.shinyapps.io/Overlap/> and making easier following the WWC *Standards* recommendations for visual analysis and also for favoring the integration of visual and statistical analysis (Horner et al., 2012). The website also reflects the fact that most quality standards include items on both visual and statistical analysis (Heyvaert, Wendt, Van Den Noortgate, & Onghena, 2015). Additionally, we have included the conservative dual criterion (Fisher et al., 2013), which is a visual aid for which there has

Running head: SINGLE-CASE LINEAR TREND

already been evidence that it improves visual analysis (Stewart, Carr, Brandt, & McHenry, 2007; Wolfe & Slocum, 2015; Young & Daly, 2016) and that it performs well in terms of avoiding false positives (Lanovaz, Huxley, & Dufour, 2017). The input of the application is a data file organized as illustrated in the webpage. The output is provided in Figures 2 and 3.

Regarding level, two lines marked with different colors are superimposed for each phase (one based on the within-phase mean and one based on the within-phase median) and the difference in means and medians is computed. Regarding trend, the MASE criterion is used to identify which of five techniques for finding a best fitting straight line (least squares, Theil-Sen, bi-split, tri-split, or differencing) leads to better fit. The slopes for the trend lines for the two phases are provided. Moreover, the values of MASE and  $R^2$  for each phase, in order to assess the goodness of fit. These values can also be understood as informing about trend stability: smaller MASE and greater  $R^2$  correspond to data that show less variability around the trend line. Note that following the MASE criterion makes unnecessary the subjective decisions regarding the straight line fitting technique to use. Moreover, fitting trend lines separately to each phase makes unnecessary the subjective decision regarding how to detrend the intervention phase data.

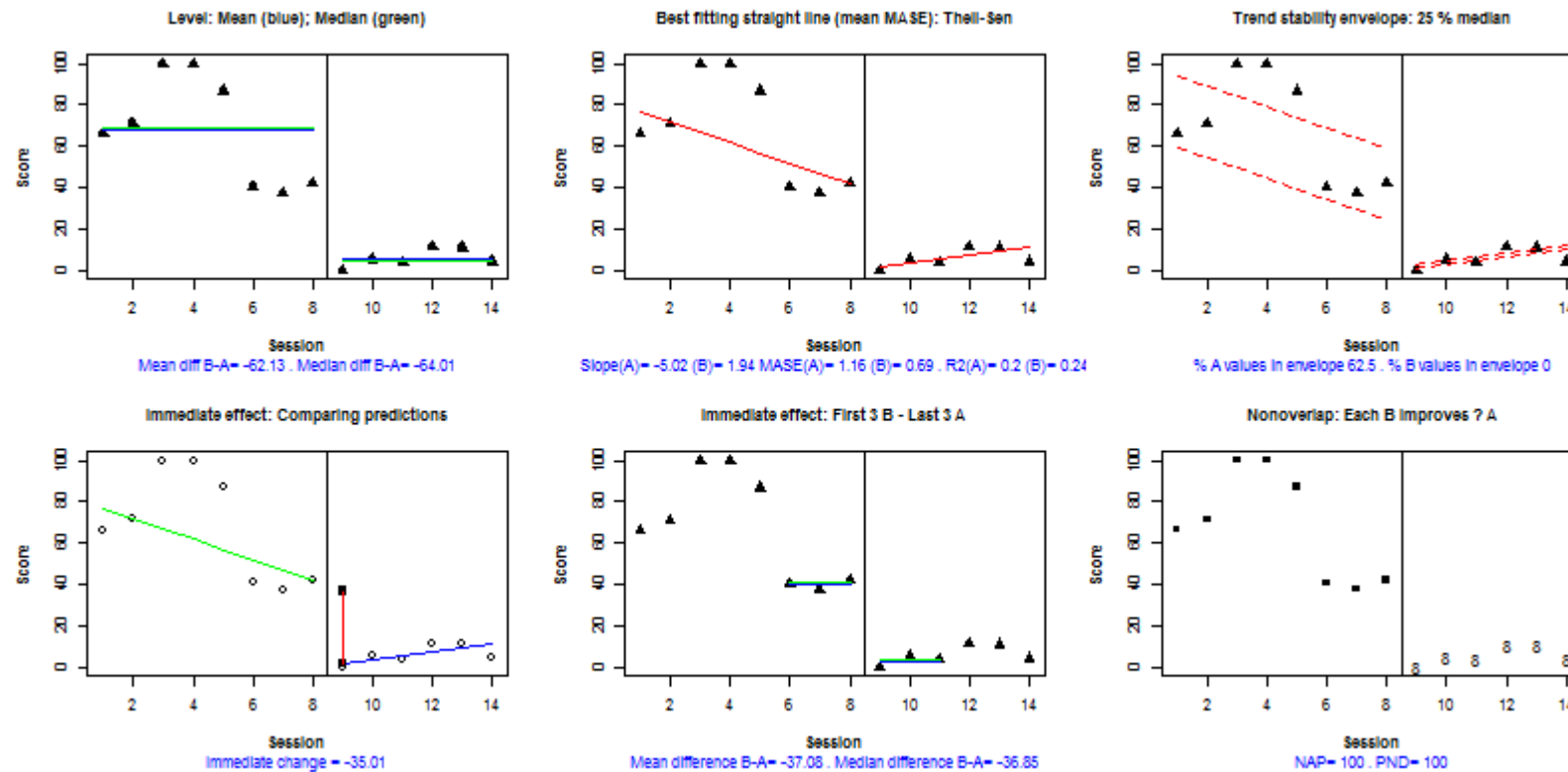
Regarding data variability, the idea of a trend stability envelope (Lane & Gast, 2014) is followed: the limits of the envelope are defined by adding and subtracting 25% of the within-phase median to the previously identified best fitting straight line. Regarding the immediacy of effect, one graphical representation and quantification focuses on the mean and median level of the last three baseline data points as compared, respectively, to the mean and median level of the first three intervention phase data points. Another graphical representation and quantification focuses on the prediction of the first

Running head: SINGLE-CASE LINEAR TREND

intervention phase data point according to the baseline vs. intervention phase trend lines, as previously identified according to MASE. Regarding overlap, the graphical representations shows, for each phase B value, the number of phase A values that it improves, apart from providing the values of the PND and the NAP.

All the aspects mentioned thus far are available in a single page (see Figure 2) with six panels, so that a visual analyst may have all this information visible simultaneously in order to respond to the “holistic and integrative nature” of visual analysis (Parker et al., 2006, p. 419). Overall, for the example data, it can be stated that the phase A data are not stable and not well represented by linear trend. Nevertheless, a general improvement is observed, especially at the end of the baseline. The phase B data are more stable and better represented by a slightly deteriorating trend line. In terms of the difference between phases, there is a clear and apparently large immediate decrease in the undesirable behavior and no overlap between the measurements of the different conditions.

Running head: SINGLE-CASE LINEAR TREND



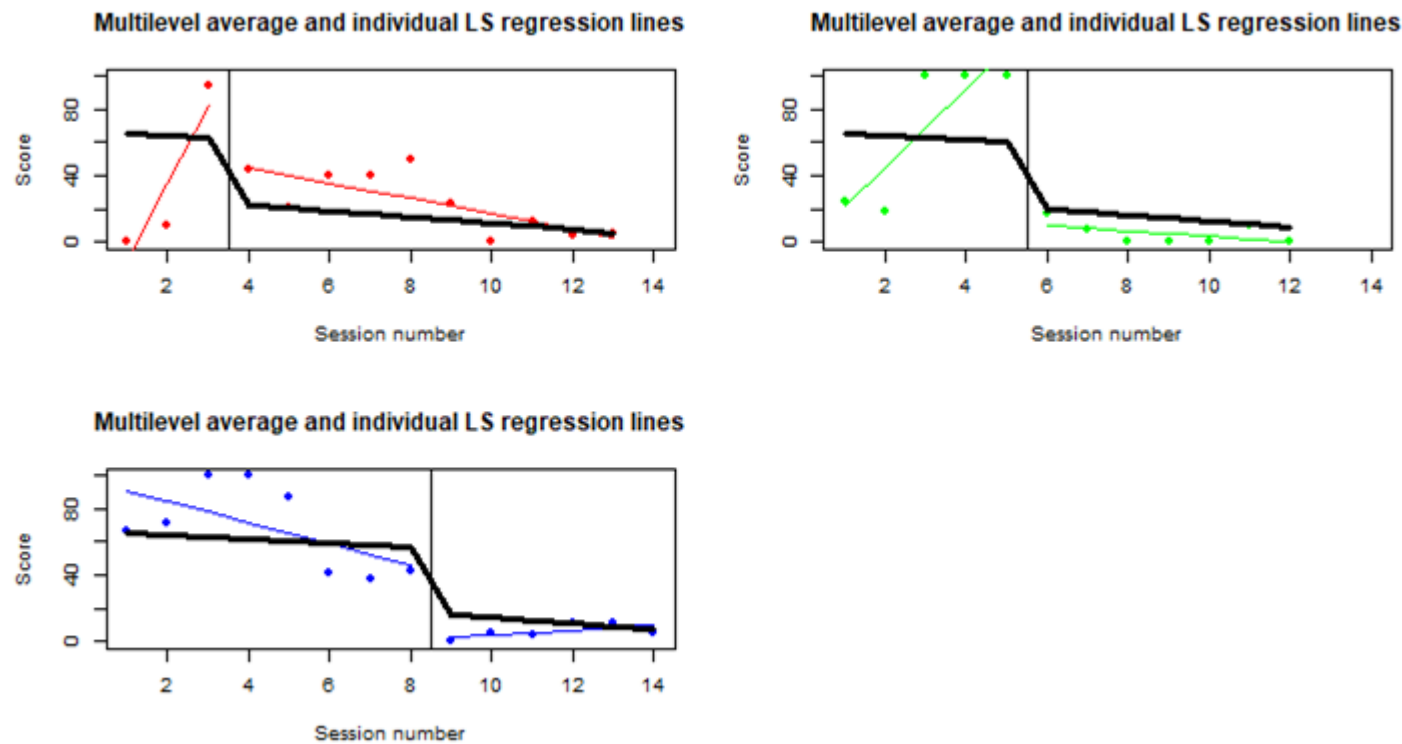
**Figure 2.** A snapshot of the web-based application for following the What Works Clearinghouse Standards recommendations for visual analysis. These data represent one tier (i.e., an AB-pair) of a nonconcurrent multiple-baseline design. The values before the vertical line refer to the baseline (A) phase, whereas the values after the vertical line refer to the intervention (B) phase

On Figure 3, we represent the way in which the assessment of consistency of data patterns is implemented in the website. As stated in the beginning of the text, inferring a causal relation between intervention and target behavior requires at least three replications and consistency in the effects. In this case, all three tiers from the nonconcurrent multiple baseline design used by Eilers and Hayes (2015) are represented in an attempt to explore whether there is a similarity between: (a) the data from the A phases; (b) the data from B phases; and (c) the type of change between phases. Specifically, a multilevel model is implemented (see Baek, Petit-Bois, Van Den Noortgate, Beretvas, & Ferron, 2016, for a visually-based example), including general trend, the effect of the intervention on the trend, and effect of the intervention on level. The average estimates obtained via the multilevel model<sup>2</sup> for the three tiers are presented with a thick line. These average estimates can be compared to the least squares regression lines fitted separately to each phase within each tier. The degree of similarity between the average levels and trends and the individual levels and trends can be used to assess the consistency of data patterns. In this particular example, there is greater similarity of the data patterns for the tiers represented in the upper part of Figure 3, whereas the tier in the lower part shows a different pattern.

---

<sup>2</sup> The multilevel model (measurements nested in individuals) is used here only as a visual aid, without dealing with the specific numerical estimates of the effects and their statistical significance.

Running head: SINGLE-CASE LINEAR TREND



**Figure 3.** A snapshot of the web-based application for following the What Works Clearinghouse Standards recommendations for visual analysis and assessing consistency. These data represent a nonconcurrent multiple-baseline design. The values before the vertical line refer to the baseline (A) phase, whereas the values after the vertical line refer to the intervention (B) phase

### **Is it Necessary to Follow the WWC *Standards*?**

It has been clearly stated by the WWC panel of experts (Kratochwill et al., 2013) that the *Standards* described in the 2010 document are pilot and not definitive. Accordingly, there have been efforts to make them more systematic by developing protocols (Maggin, Briesch, & Chafouleas, 2013) and also to emphasize the availability of additional tools related to the assessment of external validity (Hitchcock, Kratochwill, & Chezan, 2015). It is also important to state that there is a variety of rubrics proposing criteria for methodological quality (Smith, 2012) and that these rubrics do not always agree (Maggin, Briesch, Chafouleas, Ferguson, & Clark, 2014). This challenge can also be seen in the wider context of concerns expressed regarding the accuracy of WWC reports and the procedures and standards that guide them (Wood, 2017).

Despite these challenges, if the focus is put specifically on evidence standards and not on design standards, the picture looks somewhat different. The main tool for judging the availability of sufficient evidence of a functional relation is visual analysis and six data features are highlighted. These data features are well-aligned with: (a) the ones identified in empirical research (e.g., Knapp, 1983); (b) the ones discussed by authorities in the field, who were not members of the WWC panel of experts (e.g., Lane & Gast, 2014; Parsonson & Baer, 1992); (c) the data features on which the empirically-supported conservative dual criterion (Fisher et al., 2003) is based; and (d) the data features included in the set of rules used by Krueger, Rapp, Ott, Lood, and Novotny (2013) for defining what a clear change and a potential for clear change is. In that sense, it seems logical to focus on these data features, although the specific way in which they are assessed (e.g., how to estimate the best fitting trend line, how to quantify nonoverlap) and the importance of each of them when determining whether a functional relation has been established are much more open to debate. We make



Running head: SINGLE-CASE LINEAR TREND

recommendations regarding how each of the data features can be assessed, but their weight in the overall assessment of intervention effectiveness is yet to be discussed thoroughly.

Finally, assessing the cumulative amount of evidence can be done following the 5-3-20 rule (Kratochwill et al., 2013), specifying respectively the minimum number of studies, independent research teams, and participants required for considering a practice as “evidence-based”. However, this is not the only option. For instance, Lanovaz and Rapp (2016) suggest that any such recommendation should have an objective basis. In that sense, they provide criteria based on the binomial distribution for the minimum proportion of experiments with positive results in order to achieve a success rate of at least 50% and a confidence interval range of 40% or less.

### **Discussion**

In order to favor the transparency of reporting (Tate et al., 2016) and the replicability of results, we recommend that researchers clearly state exactly how a trend line is fitted and how baseline trend is controlled for, if such a control is performed. We hope that the details provided here and the information depicted in the output of the web application are helpful for transparent reporting. Additionally, similar explicit statements regarding the quantification of the difference in level, overlap, immediacy of effect or variability are also required.

### **Implications**

The main implication of the current work for applied researchers and practitioners is the possibility to carry out, online and for free, systematic visual analysis, using visual aids and quantifications referring to the six data features highlighted in the WWC *Standards*. Regarding the analytical challenges identified in the article, the question of exactly how

Running head: SINGLE-CASE LINEAR TREND

to detrend (i.e., how to remove baseline trend) is avoided by fitting trend lines separately to each phase, whereas the question about the trend line fitting technique is answered objectively by using the MASE. The quantification, implemented in the website, of the remaining data aspects has also been explained, in order to improve transparency.

The main contribution for methodologists and statisticians is the systematization of best fitting straight line techniques and detrending procedures. The formulaic representations included in the Appendices are useful for simulation studies and further software implementations in order to improve reproducibility.

### **Limitations**

Regarding the overview of techniques for finding a best fitting straight line, it was already underlined that it should not be considered comprehensive, as the focus is on techniques included in single-case data analytical procedures. Another limitation already mentioned is the focus on linear trend, which should be understood as a parsimonious simplification of reality and not as an assumption that all trends are linear. Moreover, we did not perform a simulation study to compare the different trend line fitting techniques; we rather implemented, in the web-based application, an objective criterion based on MASE for selecting the trend line fitting technique to use for each separate actually obtained data set.

Regarding the application developed, we do not claim that it includes all possible ways of implementing the WWC *Standards* recommendations for visual analysis. On the one hand, some researchers may prefer other ways of representing variability. For instance, in the SCDA [single-case data analysis] plug-in (Bulté & Onghena, 2012) for R-Commander (which is part of R; <https://www.R-project.org/>), variability can be

Running head: SINGLE-CASE LINEAR TREND

represented using range lines, range bars, and a trended range. On the other hand, the SCDA, unlike the application presented in the current text, allows representing nonlinear trends via running medians (Tukey, 1977), without assuming any particular model for the data. Thus, we encourage applied researchers to also use the SCDA plugin for visual analysis.

### 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.
- Allison, D. B., & Gorman, B. S. (1994). ‘Make things as simple as possible, but no simpler’. A rejoinder to Scruggs and Mastropieri. *Behaviour Research and Therapy*, 32, 885–890.
- Anderson, C., & Schumacker, R. E. (2003). A comparison of five robust regression methods with ordinary least squares regression: Relative efficiency, bias, and test of the null hypothesis. *Understanding Statistics*, 2, 70–103.
- Baek, E. K., Petit-Bois, M., Van Den Noortgate, W., Beretvas, S. N., & Ferron, J. M. (2016). Using visual analysis to evaluate and refine multilevel models of single-case studies. *The Journal of Special Education*, 50, 18–26.
- Bailey, D. B. (1984). Effects of lines of progress and semilogarithmic charts on ratings of charted data. *Journal of Applied Behavior Analysis*, 17, 359–365.
- Brossart, D. F., Parker, R. I., & Castillo, L. G. (2011). Robust regression for single-case data analysis: How can it help? *Behavior Research Methods*, 43, 710–719.
- Bulté, I., & Onghena, P. (2012). When the truth hits you between the eyes: A software tool for the visual analysis of single-case experimental data. *Methodology*, 8, 104–114.
- Campbell, J. M. (2013). Commentary on PND at 25. *Remedial and Special Education*, 34, 20–25.

Running head: SINGLE-CASE LINEAR TREND

- Campbell, J. M., & Herzinger, C. V. (2010). Statistics and single subject research methodology. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 417–453). London, UK: Routledge.
- Carr, J. E. (2005). Recommendations for reporting multiple-baseline designs across participants. *Behavioral Interventions*, 20, 219–224.
- 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.
- Chatfield, C. (2000). *Time-series forecasting*. London, UK: Chapman & Hall/CRC.
- Eilers, H. J., & Hayes, S. C. (2015). Exposure and response prevention therapy with cognitive defusion exercises to reduce repetitive and restrictive behaviors displayed by children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 19, 18–31.
- 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.
- Ferron, J. M., Moeyaert, M., Van Den Noortgate, W., & Beretvas, S. N. (2014). Estimating causal effects from multiple-baseline studies: Implications for design and analysis. *Psychological Methods*, 19, 493–510.
- Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis*, 36, 387–406.

Running head: SINGLE-CASE LINEAR TREND

Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd Ed.).

London, UK: Sage.

Gage, N. A., & Lewis, T. J. (2013). Analysis of effect for single-case design research.

*Journal of Applied Sport Psychology*, 25, 46–60.

Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data.

*Behavioral Assessment*, 5, 141–154.

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.

Heyvaert, M., Wendt, O., Van Den Noortgate, W., & Onghena, P. (2015).

Randomization and data-analysis items in quality standards for single-case experimental studies. *The Journal of Special Education*, 49, 146–156.

Hitchcock, J. H., Kratochwill, T. R., & Chezan, L. C. (2015). What Works

Clearinghouse standards and generalization of single-case design evidence. *Journal of Behavioral Education*, 24, 459–469.

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.

Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations

for the systematic analysis and use of single-case research. *Education and Treatment of Children*, 35, 269–290.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast

accuracy. *International Journal of Forecasting*, 22, 679–688.

Running head: SINGLE-CASE LINEAR TREND

Knapp, T. J. (1983). Behavioral analysts' visual appraisal of behavior change in graphic display. *Behavioral Assessment*, 5, 155–164.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single case designs technical documentation*. In What Works Clearinghouse: Procedures and standards handbook (Version 1.0). Available at [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_scd.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_scd.pdf)

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34, 26–38.

Kratochwill, T. R., Levin, J. R., Horner, R. H., & Swoboda, C. M. (2014). Visual analysis of single-case intervention research: Conceptual and methodological issues. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 91–125). Washington, DC: American Psychological Association.

Krueger, T. K., Rapp, J. T., Ott, L. M., Lood, E. A., & Novotny, M. A. (2013). False positives in A-B designs: Potential implications for practitioners. *Behavior Modification*, 37, 615–630.

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation*, 24, 445–463.

Lanovaz, M. J., Huxley, S. C., & Dufour, M. M. (2017, May 17). Using the dual-criteria methods to supplement visual inspection: An analysis of non-simulated data.

*Journal of Applied Behavior Analysis*. Advance online publication. doi:  
10.1002/jaba.394

Lanovaz, M. J., & Rapp, J. T. (2016). Using single-case experiments to support evidence-based decisions: How much is enough? *Behavior Modification*, *40*, 377–395.

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single-subject research: Synthesis of the self-management literature base. *Remedial and Special Education*, *34*, 44–58.

Maggin, D. M., Briesch, A. M., Chafouleas, S. M., Ferguson, T. D., & Clark, C. (2014). A comparison of rubrics for identifying empirically supported practices with single-case research. *Journal of Behavioral Education*, *23*, 287–311.

Manolov, R., Arnau, J., Solanas, A., & Bono, R. (2010). Regression-based techniques for statistical decision making in single-case designs. *Psicothema*, *22*, 1026–1032.

Manolov, R., & Moeyaert, M. (2017). How can single-case data be analyzed? Software resources, tutorial, and reflections on analysis. *Behavior Modification*, *41*, 179–228.

Manolov, R., & Rochat, L. (2015). Further developments in summarising and meta-analysing single-case data: An illustration with neurobehavioural interventions in acquired brain injury. *Neuropsychological Rehabilitation*, *25*, 637–662.

Manolov, R., & Solanas, A. (2008). Comparing  $N = 1$  effect size indices in presence of autocorrelation. *Behavior Modification*, *32*, 860–875.



- Manolov, R., & Solanas, A. (2013). A comparison of mean phase difference and generalized least squares for analyzing single-case data. *Journal of School Psychology, 51*, 201–215.
- Mercer, S. H., & Sterling, H. E. (2012). The impact of baseline trend control on visual analysis of single-case data. *Journal of School Psychology, 50*, 403–419.
- Michiels, B., Heyvaert, M., Meulders, A., & Onghena, P. (2017). Confidence intervals for single-case effect size measures based on randomization test inversion. *Behavior Research Methods, 49*, 363–381.
- Miller, M. J. (1985). Analyzing client change graphically. *Journal of Counseling and Development, 63*, 491–494.
- Novotny, M. A., Sharp, K. J., Rapp, J. T., Jelinski, J. D., Lood, E. A., & Steffes, A. K. (2014). False positives with visual analysis for nonconcurrent multiple baseline designs and ABAB designs: Preliminary findings. *Research in Autism Spectrum Disorders, 8*, 933–943.
- Ottenbacher, K. J. (1990). When is a picture worth a thousand p values? A comparison of visual and quantitative methods to analyze single subject data. *The Journal of Special Education, 23*, 436–449.
- 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., Cryer, J., & Byrns, G. (2006). Controlling baseline trend in single-case research. *School Psychology Quarterly, 21*, 418–443.

Running head: SINGLE-CASE LINEAR TREND

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., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35, 303–322.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2014). A simple method to control positive baseline trend within data nonoverlap. *The Journal of Special Education*, 48, 79–91.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, 42, 284–299.

Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current research into the stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research designs and analysis: New directions for psychology and education* (pp. 15–40). Hillsdale, NJ: Lawrence Erlbaum.

Pfadt, A., & Wheeler, D. J. (1995). Using statistical process control to make data-based clinical decisions. *Journal of Applied Behavior Analysis*, 28, 349–370.

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, 39, 368–393.

Schlosser, R. W., Lee, D. L., & Wendt, O. (2008). Application of the percentage of non-overlapping data (PND) in systematic reviews and meta-analyses: A systematic review of reporting characteristics. *Evidence-Based Communication Assessment and Intervention*, 2, 163–187.

Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education, 34*, 9–19.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*, 123–147.

Skiba, R., Deno, S., Marston, D., & Casey, A. (1989). Influence of trend estimation and subject familiarity on practitioners judgements of intervention effectiveness. *The Journal of Special Education, 22*, 433–446.

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods, 17*, 510–550.

Solanas, A., Manolov, R., & Onghena, P. (2010). Estimating slope and level change in N=1 designs. *Behavior Modification, 34*, 195–218.

Solmi, F., Onghena, P., Salmaso, L., & Bulté, I. (2014). A permutation solution to test for treatment effects in alternation design single-case experiments. *Communications in Statistics - Simulation and Computation, 43*, 1094–1111.

Solomon, B. G. (2014). Violations of assumptions in school-based single-case data: Implications for the selection and interpretation of effect sizes. *Behavior Modification, 38*, 477–496.

Stewart, K. K., Carr, J. E., Brandt, C. W., & McHenry, M. M. (2007). An evaluation of the conservative dual-criterion method for teaching university students to visually inspect AB-design graphs. *Journal of Applied Behavior Analysis, 40*, 713–718.

Running head: SINGLE-CASE LINEAR TREND

- Sullivan, K. J., Shadish, W. R., & Steiner, P. M. (2015). An introduction to modeling longitudinal data with generalized additive models: Applications to single-case designs. *Psychological Methods, 20*, 26–42.
- Swaminathan, H., Rogers, H. J., Horner, R., Sugai, G., & Smolkowski, K. (2014). Regression models for the analysis of single case designs. *Neuropsychological Rehabilitation, 24*, 554–571.
- Tarlow, K. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected Tau. *Behavior Modification, 41*, 427–467.
- Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., ..., Vohra, S. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology, 4*, 10–31.
- Tukey, J. W. (1977). *Exploratory data analysis*. London, UK: Addison-Wesley.
- 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.
- Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). London, UK: Academic Press.
- Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education, 44*, 18–29.
- Wolfe, K., & Slocum, T. A. (2015). A comparison of two approaches to training visual analysis of AB graphs. *Journal of Applied Behavior Analysis, 48*, 472–477.

Running head: SINGLE-CASE LINEAR TREND

Wood, T. W. (2017). Does the What Works Clearinghouse really work? Investigations into issues of policy, practice, and transparency. A National Institute for Direct Instruction White Paper, Office of Research and Evaluation. Retrieved on May 2, 2017 from <https://www.nifdi.org/docman/research/white-papers/1431-does-the-what-works-clearinghouse-really-work-investigations-into-issues-of-policy-practice-and-transparency/file>

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15, 642–656.

Young, N. D., & Daly III, E. J. (2016). An evaluation of prompting and reinforcement for training visual analysis skills. *Journal of Behavioral Education*, 25, 95-119.

**Appendices for  
"Linear trend in single-case visual and  
quantitative analyses"**

**(article published in Behavior Modification)**

Rumen Manolov

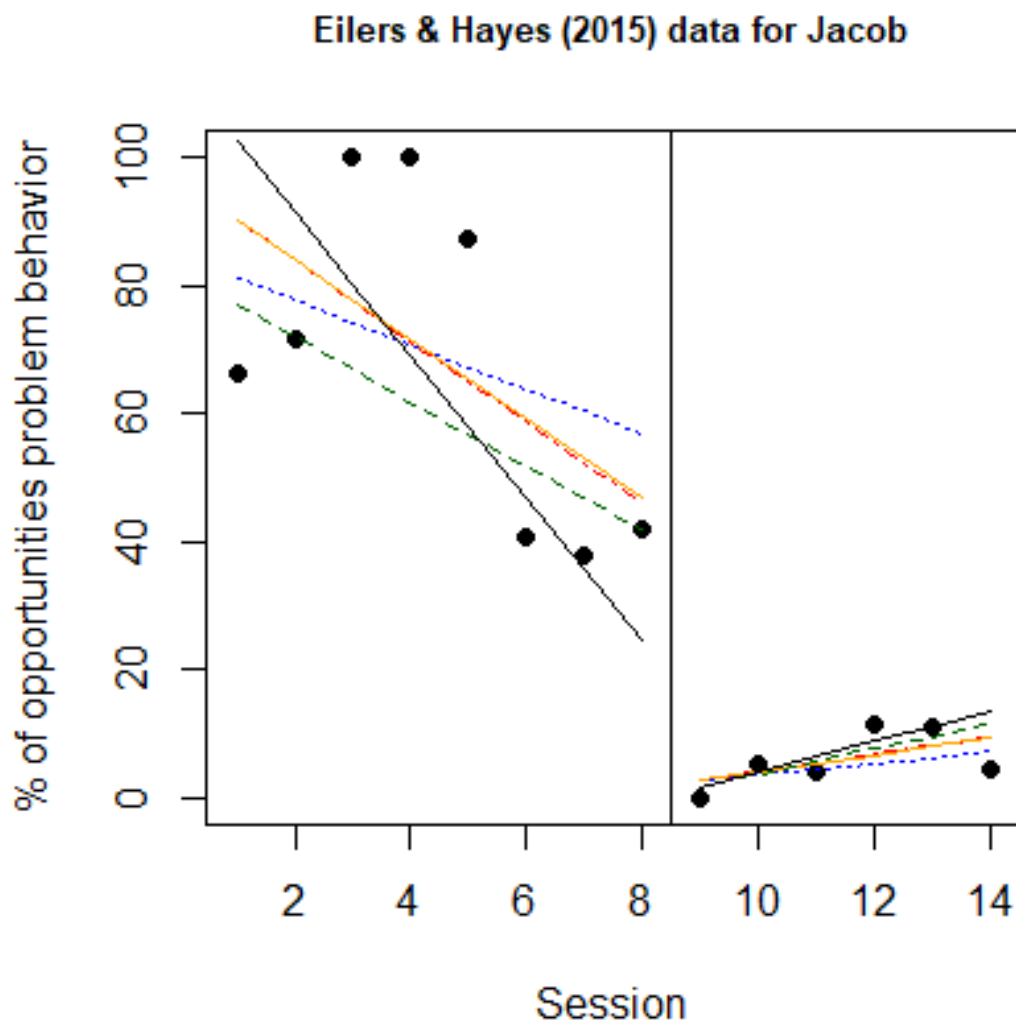
# Contents

<b>1</b>	<b>Presenting the example</b>	<b>2</b>
<b>2</b>	<b>Appendix A: What trend line fitting techniques exist and how do they differ?</b>	<b>3</b>
2.1	Notation . . . . .	3
2.2	Straight line fitting techniques formally presented . . . . .	4
2.2.1	Least squares . . . . .	4
2.2.2	Bi-split . . . . .	5
2.2.3	Tri-split . . . . .	6
2.2.4	Theil-Sen estimator . . . . .	8
2.2.5	Differencing . . . . .	9
2.2.6	Summary table of linear trends . . . . .	10
2.3	A quantification of monotonic trend . . . . .	11
<b>3</b>	<b>Appendix B: Comparing trend lines via measures of fit</b>	<b>12</b>
3.1	Well-known options . . . . .	12
3.2	Our recommendation . . . . .	12
<b>4</b>	<b>Appendix C: What ways of controlling for baseline trend have been incorporated in single-case data analytical procedures?</b>	<b>13</b>
4.1	Differencing analysis . . . . .	13
4.2	Trend analysis . . . . .	14
4.3	Allison and Gorman (1993) Regression Model . . . . .	15
4.4	Mean and Slope Adjustment (MASAJ) . . . . .	16
4.5	Tau-U . . . . .	17
4.6	Baseline Corrected Tau . . . . .	18
4.7	Detrending in the Between-Case Standardized Mean Difference (BC-SMD) . . . . .	19
4.8	Slope and Level Change (SLC) . . . . .	20
4.8.1	Summary table of results after detrending . . . . .	21
<b>5</b>	<b>References</b>	<b>22</b>

# 1 Presenting the example

The data we use as a main illustration are collected from a participant called Jacob, who is part of the multiple-baseline data presented by Eilers and Hayes (2015; Experiment 1) on problematic (repetitive and restrictive) behavior in children with autism spectrum disorder, with the intervention consisting of cognitive defusion exercise and exposure therapy. The data were retrieved using Plot Digitizer for Windows (plotdigitizer.sourceforge.net), obtaining the following values: 66.39, 71.55, 100, 100, 87.06, 40.74, 37.7, 42.12, 0, 5.44, 3.89, 11.31, 11.25, and 4.48.

Figure 1 provides a graphical representation of the data. Black continuous line: bi-split technique for estimating trend; Orange solid line: tri-split technique for estimating trend. Dash-and-dot red line: ordinary least squares. Blue dotted line: differencing. Dark green dashed line: Theil-Sen. The values before the vertical line refer to the baseline (A) phase, whereas the values after the vertical line refer to the intervention (B) phase.





## 2 Appendix A: What trend line fitting techniques exist and how do they differ?

### 2.1 Notation

In the formulas that we present in the current section, we assume that a trend line is fitted to the data of a single phase. The number of measurements in the phase would be denoted by  $n$ , which can be denoted by  $n_A$  when referring to the baseline data or by  $n_B$  when referring to the intervention phase data. The measurement occasions will be denoted by  $t$ , including the ordinal numbers  $1, 2, \dots, n$ . The actually obtained values of the outcome variable will be denoted by  $y_t$ , indicating the measurement occasion  $t$  to which they refer. The mean of the actually obtained values is denoted by  $\bar{y}$ . The fitted values obtained by any of the trend fitting techniques will be denoted by  $\hat{y}_t$ . The median of a set of values will be represented as  $Md()$ , with the values being specified in parenthesis. Additional terms  $(g, h, i, j, k, l, r, s, u)$  will be defined in each of the expressions, whenever necessary. Several letters are used in order to avoid using two different definitions for the same term.

## 2.2 Straight line fitting techniques formally presented

### 2.2.1 Least squares

First, regarding ordinary least squares estimation, the trend line is fitted in such a way that the squared difference between itself and the actual values is the smallest possible. The least squares estimate of the slope of the baseline trend line is obtained as:

$$b_{1(LS)} = \frac{\sum_{t=1}^n (y_t - \bar{y})(t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2} \quad (1)$$

The intercept is estimated as:

$$b_{0(LS)} = \bar{y} - b_{1(LS)} \times \bar{t} \quad (2)$$

Using the phase A data from the example represented on Figure 1,  $\bar{y} = \frac{(66.39+71.55+100+100+87.06+40.74+37.7+42.1)}{8} = 545.56/8 = 68.195$  and  $\bar{t} = \frac{1+2+3+4+5+6+7+8}{8} = \frac{36}{8} = 4.5$ .  $b_{1(LS)} = \frac{\sum_{t=1}^n (y_t - \bar{y})(t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2} = -6.308$  and  $b_{0(LS)} = \bar{y} - b_{1(LS)} \times \bar{t} = 68.195 - (-6.308) \times 4.5 = 96.581$ .

### 2.2.2 Bi-split

Bi-split (Parker et al., 2014) has also been referred to as the split-middle method (Gast & Spriggs, 2010; Miller, 1985) and extended celeration line (O. R. White & Haring, 1980). It entails splitting the data into two equally-sized sections and connecting the medians of these sections with a straight line. Its main value is its simplicity and possibility for hand calculation (Miller, 1985).

When  $n$  is an even number, each of the portions contains  $n/2$  data points. When  $n$  is an odd number the dividing line falls on the middle measurement occasion (Miller, 1985) and thus the middle value will not be used when defining the medians (Lane & Gast, 2014). In the expressions we present here we follow this logic, which is also similar to the one followed when defining the tri-split (Velleman & Johnstone, 1985). A different approach is followed in the SCDA plug-in for R (Bulté & Onghena, 2012) in which two different split-middle lines are represented according to the portion of data to which the extra measurement is assigned.

Formally the slope is defined as

$$b_{1(bi)} = \frac{Md(y_j, \dots, y_n) - Md(y_1, \dots, y_i)}{\frac{j+n}{2} - \frac{1+i}{2}} \quad (3)$$

with

$$n \bmod 2 = \begin{cases} 0: & i = n/2; j = (n/2) + 1 \\ 1: & i = ((n-1))/2; j = ((n+1))/2 + 1 \end{cases}$$

In the Equation 3  $i$  is the last value included in the first portion of data and  $j$  is the first value included in the second portion;  $n \bmod 2$  refers to the remainder of the division  $n/2$ . The denominator of  $b_{1(bi)}$  includes the following expressions for the middle points of the two portions:  $Md(1, 2, \dots, i) = ((1+i))/2$  and  $Md(j, j+1, \dots, n) = ((j+n))/2$ .

Regarding the intercept, it is obtained as

$$b_{0(bi)} = Md(y_1, y_2, \dots, y_i) - b_{1(bi)} \times \frac{1+i}{2} \quad (4)$$

Finally, it is possible to shift the trend line upwards or downwards until 50% of the data are above the line and 50% are below it.

Using the phase A data from the example represented on Figure 1,  $n_A = 8$  and thus  $n_A \bmod 2 = 0$ . Therefore, the last measurement in the first portion is  $i = n/(2 = 8/2 = 4)$  and the first measurement in the second portion is  $j = (n/2) + 1 = (8/2) + 1 = 5$ .  $Md(y_j, y(j+1), \dots, y_n) = Md(y_5, y_6, y_7, y_8) = Md(87.06, 40.74, 37.7, 42.12) = ((40.74 + 42.12))/2 = 41.43$ .  $Md(y_1, y_2, \dots, y_i) = Md(y_1, y_2, y_3, y_4) = Md(66.39, 71.55, 100, 100) = ((71.55 + 100))/2 = 85.775$ . With these values, the slope and the intercept are estimated, respectively, as follows:

$$b_{1(bi)} = \frac{41.43 - 85.775}{\frac{5+8}{2} - \frac{1+4}{2}} = \frac{-44.345}{6.5 - 2.5} \simeq -11.09$$

$$b_{0(bi)} = 85.775 - (b_{1(bi)}) - 11.09 \times \frac{1+4}{2} = 113.5$$

### 2.2.3 Tri-split

In tri-split (also known as three-group resistant line, Johnstone Velleman, 1985), the data of a phase are divided into three equally-sized portions and the medians of the first and third portions are used for estimating the trend, whereas as the medians of all three portions are used for estimating the intercept. The aim is for the tri-split trend line to be resistant to outlying values (Velleman & Hoaglin, 1981).

Regarding how exactly the split is performed, Vannest, Davis, and Parker (2013) mention that the usual recommendation (e.g., Bartlett, 1949) is for the three slices or portions of data to be of the same size (when the focus is on the baseline, each portion would be  $n_A/3$ ), but they suggest an additional rule with the following size of the three portions: 27%, 46%, and 27%. Vannest et al. (2013) base their recommendation on McCabe's (1980) article, but McCabe alerts that the favorable results for the 27% rule (or, even better, the 25% rule which is equivalent to the 1:2:1 ratio) is based on asymptotic results and assumptions of normality. Accordingly, Gibson and Jowett (1957) also suggest the 1:2:1 rule (having 25% of the data in the first portion, 50% in the second portion, and 25% in the third portion) for a normal independent variable (X), but they suggest the ratio 1:1:1 (i.e., equally-size portions) when X is uniform, as is the case for the measurement occasions (each of the value  $1, 2, \dots, n_A$  occurs the same number of times: once). Velleman and Hoaglin (1981) recommend the following rules for achieving a split into approximately equal slices when  $n_A$  is not a multiple of 3: (a) when the remainder of the division  $n_A/3$  is 1 (e.g., when  $n_A = 7$ ), the exceeding value goes into the middle portion of the data (two values in portion 1, three values in portion 2, and two values in portion 3); and (b) when the remainder of the division  $n_A/3$  is 2 (e.g., when  $n_A = 8$ ), the exceeding values go into the extreme portions of the data (three values in portion 1, two values in portion 2, and three values in portion 3).

Thus, the slope is estimated formally as

$$b_{1(tri)} = \frac{Md(y_k, \dots, y_n) - Md(y_1, \dots, y_l)}{\frac{k+n}{2} - \frac{1+l}{2}} \quad (5)$$

with

$$n \bmod 3 = \begin{cases} 0: & l = n/3; k = n - (n/3) + 1 \\ 1: & l = \lfloor U + 230A \rfloor / n/3; k = n - \lfloor n/3 \rfloor < U + 230B > + 1 \\ 2: & l = \lfloor U + 230A \rfloor / n/3; k = n - \lfloor n/3 \rfloor \end{cases}$$

In the Equation 5,  $l$  is the last value included in the first portion of data and  $k$  is the first value included in the third portion;  $n \bmod 3$  refers to the remainder of the division  $n/3$ . The denominator of  $b_{1(tri)}$  includes the following expressions for the middle points of the two portions:  $Md(1, 2, \dots, l) = ((1+l))/2$  and  $Md(k, k+1, \dots, n) = ((k+n))/2$ . For estimating the intercept Velleman and Hoaglin (1981) recommend using the average of the intercepts estimated from each of the three portions of data, so that the intercept is not determined only by the middle portion:

$$\begin{aligned} b_{0(first)} &= Md(y_1, y_2, \dots, y_l) - b_{1(tri)} \times \frac{1+l}{2} \\ b_{0(second)} &= Md(y_{l+1}, y_{l+2}, \dots, y_{k-1}) - b_{1(tri)} \times \frac{(1+l) + (k-1)}{2} \\ b_{0(third)} &= Md(y_k, y_{k+1}, \dots, y_n) - b_{1(tri)} \times \frac{k+n}{2} \\ b_{0(tri)} &= \frac{b_{0(first)} + b_{0(second)} + b_{0(third)}}{3} \end{aligned} \quad (6)$$

Using the phase A data from the example represented on Figure 1,  $n_A = 8$  and thus  $n_A \bmod 3 = 2$ . Therefore, the last measurement in the first portion is

$$l = \lfloor 8/3 \rfloor + 1 = \lfloor 2.67 \rfloor < U + 230B > + 1 = 2 + 1 = 3$$

and the first measurement in the third portion

$$k = 8 - \lfloor 8/3 \rfloor = 8 - \lfloor 2.67 \rfloor = 8 - 2 = 6$$

$Md(y_k, y_{k+1}, \dots, y_n) = Md(y_6, y_7, y_8) = Md(40.74, 37.7, 42.12) = 40.74$ .  $Md(y_1, y_2, \dots, y_l) = Md(y_1, y_2, y_3) = Md(66.39, 71.55, 100) = 71.55$ . With these values, the slope is estimated follows:

$$b_{1(tri)} = \frac{40.74 - 71.55}{\frac{6+8}{2} - \frac{1+3}{2}} = \frac{-30.81}{7-2} = -6.162$$

The intercept is estimated follows:

$$b_{0(first)} = 71.55 - (-6.162) \times \frac{1+3}{2} = 83.874$$

$$\begin{aligned} b_{0(second)} &= Md(y_4, y_5, \dots, y_{k-1}) - (-6.162) \times \frac{(1+l) + (k-1)}{2} \\ &= Md(87.06, 100) + 6.612 \times \frac{(4) + (5)}{2} \\ &= 93.53 + 27.729 \\ &\simeq 121.26 \end{aligned}$$

$$b_{0(third)} = 40.74 - (-6.162) \times \frac{6+8}{2} = 83.874$$

$$b_{0(tri)} = \frac{b_{0(first)} + b_{0(second)} + b_{0(third)}}{3} = \frac{83.874 + 121.26 + 83.874}{3} \simeq 96.34$$

#### 2.2.4 Theil-Sen estimator

The Theil-Sen estimator can be understood as the “median slope of many ‘mini-slopes’ created from all pairwise data comparisons made in time order (early to late) in a time series” (Parker et al., 2014, p. 81). Another way of conceptualizing the Theil-Sen estimator is as a way of estimating slope,  $b$ , so that the Kendall’s correlation between the measurement occasions,  $t$ , and the difference  $(y_t - b \times t)$  is approximately equal to zero (Wilcox, 2012). The Theil-Sen estimator has been presented (Sen, 1968) as a robust alternative to the least squares estimator in case of outliers and non-normality. The Theil-Sen estimator (Sen, 1968; Theil, 1950) estimates slope as:

$$b_{1(TS)} = Md\left(\frac{y_r - y_s}{r - s}\right); r = 1, 2, \dots, t; s = 1, 2, \dots, n; n \neq s \quad (7)$$

The intercept is estimated as the median of the differences between the actual measurements and the values predicted according to the slope and the measurement occasion. Formally,

$$b_{0(TS)} = Md(y_t - b_{1(TS)}t) \quad (8)$$

Using the phase A data from the example represented on Figure 1, there are  $n_A(n_A - 1) = 8 \times 7 = 56$  pairs of data points and thus 56 slopes are computed. The median of these slopes is  $b_{1(TS)} = -5.0175$ . For computing the intercept, it is necessary to obtain  $y_1 - b_{1(TS)} \times 1 = 66.39 - (-5.0175)1 = 71.4075$ ,  $y_2 - b_{1(TS)}1 = 71.55 - (-5.0175) \times 2 = 81.585$ , up to  $y_8 - b_{1(TS)}8 = 42.12 - (-5.0175) \times 8 = 82.26$ . The median of these eight values is  $b_{0(TS)} = 81.9225$ .

### 2.2.5 Differencing

Differencing refers to the operation of subtracting each data point from the next one and thus the  $n - 1$  differenced values are obtained as  $y'_t = y_{t+1} - y_t$ . Differencing is used here for estimating and not for removing trend (Solanas, Manolov, & Onghena, 2010). In analytical procedures using differencing for fitting a straight line, the slope represents the average of the differences between successive values. Formally,

$$b_{1(D)} = \bar{y}'_t = \frac{\sum_{t=1}^{n-1} y'_t}{n-1} \quad (9)$$

Regarding the estimate of intercept, it is necessary to refer to a single-case data analytical procedure called the Mean Phase Difference. In the initial version (Manolov Solanas, 2013; hereinafter  $MPD_{2013}$ ) baseline trend is extrapolated into the intervention phase by adding the trend estimate ( $b_{1(D)}$ ) to the last baseline phase data point ( $y_{n_A}$ ) so that the first fitted intervention phase value is  $\hat{y}_{n_A+1} = y_{n_A} + b_{1(D)}$ . From this specification, it can be derived that the intercept is defined formally, using the general term  $n$  instead of  $n_A$  as:

$$b_{0(MPD_{2013})} = y_n - b_{1(D)} \quad (10)$$

In the modified version of the procedure (Manolov Rochat, 2015; hereinafter  $MPD_{2015}$ ) it was suggested that the trend line should pass through the point that corresponds to the median measurement occasion  $Md(t) = ((n+1)/2)$  in the phase and the median value  $Md(y_t)$ . From this specification, it can be derived that the intercept is formally defined as:

$$b_{0(MPD_{2015})} = Md(y_t) - b_{1(D)} \frac{n+1}{2} \quad (11)$$

Using the phase A data from the example represented on Figure 1, it is first necessary to obtain the differenced values:  $y'_1 = y_2 - y_1 = 71.55 - 66.39 = 5.16$ ;  $y'_2 = y_3 - y_2 = 100 - 71.55 = 28.45$  up to  $y'_7 = y_8 - y_7 = 42.12 - 37.70 = 4.42$ . The average of the seven differenced values is

$$b_{1(D)} = \frac{5.16 + 28.45 + 0 + (-12.94) + (-46.32) + (-3.04) + 4.42}{7} \simeq -3.47$$

In Figure 1,  $b_{0(MPD_{2015})}$  is used for defining the intercept and thus it is necessary to obtain the median of all eight baseline phase value  $Md(y_t) = Md(66.39, 71.55, 100, 100, 87.06, 40.74, 37.7, 42.12) = \frac{66.39+71.55}{2} = 68.97$ .

$$b_{0(MPD_{2015})} = Md(y_t) - b_{1(D)} \frac{n+1}{2} = 68.97 - (-3.47) \frac{8+1}{2} \simeq 84.57$$

In case  $b_{0(MPD_{2013})}$  had been used, the intercept would have been

$$b_{0(MPD_{2013})} = y_n - b_{1(D)} = 42.12 - (-3.47) \times 8 = 68.88$$

### 2.2.6 Summary table of linear trends

As a summary, Table A.1 includes the estimates of intercept and slope according to the different techniques reviewed in the current Appendix A.

**Table A.1.** Intercept and slope estimates for the best fitting straight line to the two phases depicted in Figure 1, according to several techniques.

Technique	Baseline phase		Intervention phase	
	Intercept	Slope	Intercept	Slope
Least squares	96.581	-6.308	1.337	1.350
Theil-Sen	81.822	-5.018	-0.177	1.937
Bi-split	113.490	-11.087	-1.017	2.453
Tri-split	96.336	-6.162	1.560	1.286
Differencing	84.572	-3467	1.824	0.896

For Differencing the definition of the intercept is according to Equation (11) from Appendix A



### 2.3 A quantification of monotonic trend

The criteria used for selecting the trend line fitting techniques on which to focus were their inclusion in single-case data analytical procedures and the fact that a straight line is being fit. There is an analytical procedure called Tau-U (Parker, Vannest, Sauber, & Davis, 2011), which also takes baseline trend into account, but it does not entail estimating linear trend; monotonic trend is rather quantified. We decided to include the technique for quantifying monotonic trend that is incorporated in Tau-U for three reasons: (a) the main text makes reference to Baseline corrected Tau (Tarlow, 2017), which is a modification of Tau-U; (b) to the best of our knowledge, the formula for quantifying and correcting monotonic trend in Tau-U is not available elsewhere (e.g., Parker et al., 2011, only provide numerical examples); and (c) we later comment on how baseline trend is controlled for in Tau-U, which requires explaining first how this trend is quantified.

Specifically, Parker et al. (2011) propose to quantify the amount of improvement in the baseline as the number of pairwise comparisons for which a baseline measurement is larger than a previous baseline measurements minus the number of comparisons for which a baseline measurement is smaller than a previous baseline measurements, divided by the number of pairwise comparisons; formally:

$$\tau_a = \frac{(y_d > y_c) - (y_d < y_c)}{\frac{n_A(n_A-1)}{2}}; c = 1, 2, \dots, n_A - 1; d = 2, 3, \dots, n_A; c < d \quad (12)$$

From the expression it is clear that  $y_d$  and  $y_c$  are baseline phase measurements,  $c$  and  $d$  are measurement occasions, with each value being compared only to previous measurement occasions.

For the phase A of the running example, given that  $n_A = 8$ , the number of comparisons to perform is equal to  $(n_A(n_A - 1))/2 = (8(8 - 1))/2 = 28$ . The second measurement (71.55) is higher than 1 previous measurement and lower than 0 previous measurements. The third measurement (100) is higher than 2 previous measurements and lower than 0. The fourth measurement (100) is higher than 2 previous measurements and lower than 0. The fifth measurement (87.06) is higher than 2 previous measurements and lower than 2. The sixth measurement (40.74) is higher than 0 previous measurements and lower than 5. The seventh measurement (37.7) is higher than 0 previous measurements and lower than 6. The eighth measurement (42.12) is higher than 2 previous measurements and lower than 5. Formally,

$$\begin{aligned} \tau_a &= \frac{(1 - 0) + (2 - 0) + (2 - 0) + (2 + 2) + (0 - 5) + (0 - 6) + (2 - 5)}{\frac{n_A(n_A-1)}{2}} \\ &= \frac{1 + 2 + 2 + 0 - 5 - 6 - 3}{28} = \frac{-9}{28} \simeq -0.3214 \end{aligned}$$

This value can be interpreted as approximately one third of the measurements being lower than previously obtained measurements.

### 3 Appendix B: Comparing trend lines via measures of fit

#### 3.1 Well-known options

Formally, an  $R^2$ -type of index, as a measure of fit, can be defined as the complementary to the proportion of residual variability:

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)}{\sum_{t=1}^n (y_t - \bar{y}_t)} \quad (13)$$

It will provide greatest value for least squares estimation, given that it minimizes  $\sum_{t=1}^n (y_t - \hat{y}_t)^2$ .

The coefficient of variation of the residual around the trend line, as a relative measure of lack of fit (Mendenhall & Sincich, 2012), is quantified as

$$CV = \frac{\sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}}{|\hat{y}|} \quad (14)$$

The Mean Square Error is based on a quadratic loss function and is computed as

$$\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} \quad (15)$$

The Mean Absolute Error is based on a linear loss function and is computed as

$$\frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n} \quad (16)$$

#### 3.2 Our recommendation

Our suggestion is to use the Mean Absolute Scaled Error (MASE), proposed by Hyndman and Koehler (2006). For a trend line fitted to the  $n_A$  baseline measurements, MASE can be written as follows:

$$MASE = \frac{\sum_{t=1}^{n_A} \left| \frac{(y_t - \hat{y}_t)}{\frac{\sum_{u=2}^n |y_u - y_{u-1}|}{n-1}} \right|}{n} \quad (17)$$

Hyndman and Koehler (2006, p. 687) state that MASE is “easily interpretable, because values of MASE greater than one indicate that the forecasts are worse, on average, than in-sample one-step forecasts from the naïve method”. (The naïve method entails predicting a value from the previous one, that is, the random walk model that has frequently been used to assess the degree to which more sophisticated methods provide more accurate forecasts than this simple procedure; Chatfield, 2000.) Thus, values of MASE greater than one could be indicative that a general trend (e.g., a linear one as in MPD) does not provide a good enough fit to the data from which it was estimated, because it does not improve the fit of the naïve method.

## 4 Appendix C: What ways of controlling for baseline trend have been incorporated in single-case data analytical procedures?

### 4.1 Differencing analysis

Differencing Analysis In differencing, which is part of autoregressive integrated moving average (ARIMA) models (Harrop & Velicer, 1985), each measurement is subtracted from subsequent measurements to eliminate linear trend. In one of the data analysis proposals by Gorsuch (1983), differencing analysis, a regression analysis is performed on the differenced data ( $y'_t = y_{t+1} - y_t$ ) rather than on the original measurements ( $y_t$ ). For the running example, differencing the original values would lead to  $y'_1 = y_2 - y_1 = 71.55 - 66.39 = 5.16$ ,  $y'_2 = y_3 - y_2 = 100 - 71.55 = 28.45$ , up to  $y'_{13} = y_{14} - y_{13} = 4.48 - 11.25 = -6.77$ . The main analysis is performed with the differenced data.

## 4.2 Trend analysis

Gorsuch's (1983) Trend analysis entails performing a regression analysis with all  $n = n_A + n_B$  values using the model

$$\hat{y} = b_0 + b_1t + b_2Phase \quad (18)$$

In that sense, the quantification of the difference in level (i.e.,  $b_2$ ) is computed partialling out the effect of general trend. Gorsuch (1983) also mentions the possibility to enter the predictors  $t$  and  $Phase$  hierarchically (i.e., sequentially) instead of simultaneously. For the running example, the use of the simultaneous multiple regression model from Equation (18) would result in the following estimates:  $b_0 = 86.445$ ,  $b_1 = -4.056$ , and  $b_2 = -33.745$ . In contrast, if trend were not considered and the model  $\hat{y}_t = b_0 + b_1Phase$  were used, the estimate of average change in level would be -62.13 instead of -33.745.

Another option for performing trend analysis (Parker & Brossart, 2003) is to first fit the model  $\hat{y}_t = b_0 + b_1t$  and then perform a second regression analysis on the residuals of the previous analysis:  $e_t = b_0 + b_1Phase$ . For the running example, such an analysis would lead to the following equation for the first model  $\hat{y}_t = 98.682 + (-7.615)t$ , indicating an overall decrease of more than 7.5 per measurement occasion; for the residuals the regression equation would be  $e_t = 3.782 + (-8.825)Phase$ , i.e., the estimate of the average change in level would be -8.825.

### 4.3 Allison and Gorman (1993) Regression Model

In this model a trend line is fitted only to the  $n_A$  baseline data only according to Equations (1) and (2). The following steps are: (a) the straight line fitted to the baseline is extrapolated into the intervention phase, so that the fitted values are obtained for all  $n = n_A + n_B$  values:  $\hat{y}_t = b_{0(LS)} + b_{1(LS)}t$ ; (b) the residuals for both the baseline and the intervention phase are computed as the difference between the actual  $y_t$  values and the predicted values  $\hat{y}_t$ , that is,  $e_t = y_t - \hat{y}_t$ ; (c) a regression analysis is performed for using the residuals as a dependent variable, modelling both change in level and in slope (if these are in opposite directions) or only change in level (otherwise). For the running example, we already obtained  $b_{1(LS)} = -6.308$  and  $b_{0(LS)} = 96.581$ . The predicted values for both phase are obtained as follows:  $\hat{y}_1 = 96.581 - 6.308 \times 1 \simeq 90.27$ ,  $\hat{y}_2 = 96.581 - 6.308 \simeq 83.96$ , up to  $\hat{y}_{14} = 96.581 - 6.308 \times 14 \simeq 8.27$ . The residuals are obtained as  $e_1 = y_1 - \hat{y}_1 = 66.39 - 90.27 = -23.88$ ,  $e_2 = y_2 - \hat{y}_2 = 71.55 - 83.96 = -12.41$ , up to  $e_{14} = y_{14} - \hat{y}_{14} = 4.48 - 8.27 = -3.79$ . The main analysis is performed with the residuals.

#### 4.4 Mean and Slope Adjustment (MASAJ)

Parker et al. (2006) proposed a modification of the way in which Allison and Gorman's (1993) regression model controls for baseline trend. According to MASAJ, baseline trend is removed only from the intervention phase data, maintaining the baseline data intact, while also maintaining the scale of the original scores. Thus, the steps are: (a) apply Equations (1) and (2) to estimate baseline trend intercept and slope; (b) obtain the  $n = n_A + n_B$  fitted values:  $\hat{y}_t = b_{0(LS)} + b_{1(LS)}t$ ; (c) obtain the  $n_A + n_B$  residuals via  $e_t = y_t - \hat{y}_t$ ; (d) replace the first  $n_A$  residuals by the actual measurements  $y_t$ ; and (e) add the mean of the baseline data to the last  $n_B$  residual. In one expression, the detrending for the intervention phase data can be formalized as:

$$\tilde{y} = y_t - (b_{0(LS)} + b_{1(LS)}t) + \frac{\sum_{h=1}^{n_A} y_h}{n_A}; t = n_A + 1, \dots, n_A + n_B \quad (19)$$

For the running example, the least squares estimates of the intercept and the slope of the baseline trend line are  $b_{0(LS)} = 96.581$  and  $b_{1(LS)} = -6.308$ , whereas the mean of the baseline data is 68.195. The detrended intervention phase values are  $\tilde{y}_{(8+1)} = 0 - (96.581 + (-6.308) \times (8+1)) + 68.195 = 28.386$ ,  $\tilde{y}_{(8+2)} = 5.44 - (96.581 + (-6.308) \times (8+2)) + 68.195 = 40.133$  up to  $\tilde{y}_{(8+6)} = 4.48 - (96.581 + (-6.308) \times (8+6)) + 68.195 = 64.406$ .

## 4.5 Tau-U

Among nonoverlap indices, for Tau-U (Parker et al., 2011) the way in which trend is controlled for is not the same in the Parker et al. (2011) article and in the official website for implementing Tau-U (<http://www.singlecaseresearch.org/calculators/tau-u>). According to the example provided in Parker et al. (2011) and also according to the R code mentioned in Brossart, Vannest, Davis and Patience (2014: <http://ktarlow.com/stats/r/tauu.txt> by Kevin Tarlow) Tau-U with baseline trend control (also referred to as  $Tau - U_{Avs.B-trendA}$ ) is computed :

$$\frac{[(y_g > y_h) - (y_g < y_h)] - (y_d > y_c) - (y_d < y_c)}{[n_A n_B] + [\frac{n_A(n_A-1)}{2}]}; g = 1, \dots, n_B; h = 1, \dots, n_A \quad (20)$$

where  $y_g$  and  $y_h$  are intervention phase measurements and the remaining terms are as defined previously. However, according to the way in which Tau-U Avs.B-trendA is computed in the website (<http://www.singlecaseresearch.org/calculators/tau-u>; Vannest, Parker, Gonen, & Adiguzel, 2016), the formula would be:

$$\frac{[(y_g > y_h) - (y_g < y_h)] - (y_d > y_c) - (y_d < y_c)}{[n_A n_B]} \quad (21)$$

What is common is that in both cases the number of measurements improving in the course of the baseline only (i.e., baseline measurements improving previous baseline measurements) is subtracted from the number of intervention phase measurements that improve baseline measurements (using all pairwise comparison as in NAP). However, the denominator of the resulting index is different.

For the running example, there are  $n_A n_B = 8 \times 6 = 48$  comparisons between baseline and intervention phase measurements and  $[(y_g > y_h) - (y_g < y_h)] = -48$ , because all intervention phase data are lower than the baseline data.

The results for Tau-U according to the Parker et al. (2011) illustration and the R code:

$$\frac{[(y_g > y_h) - (y_g < y_h)] - (y_d > y_c) - (y_d < y_c)}{[n_A n_B] + [\frac{n_A(n_A-1)}{2}]} = \frac{-48 - (-9)}{48 + 28} = \frac{-39}{76} \simeq -0.5132$$

The result for Tau-U according to the website is:

$$\frac{[(y_g > y_h) - (y_g < y_h)] - (y_d > y_c) - (y_d < y_c)}{[n_A n_B]} = \frac{-48 - (-9)}{48} = \frac{-39}{48} \simeq -0.8125$$

## 4.6 Baseline Corrected Tau

According to the “additional output” of <http://ktarlow.com/stats/tau/>, the Baseline corrected Tau (Tarlow, 2017) removes baseline trend from the baseline and intervention phase data using the expression

$$\tilde{y}_t = y_t - b_{1(LS)}t; t = 1, \dots, n_A + n_B \quad (22)$$

This correction entails that the original measurement units of the dependent variable are maintained.

For the running example using the data from both phases depicted in Figure 1,  $\tilde{y}_1 = 66.39 - (-5.0175) \times 1 = 71.4075$ ,  $\tilde{y}_2 = 71.55 - (-5.0175 \times 2 = 81.5850$ , up to  $\tilde{y}_{14} = 4.48 - (-5.0175) \times 14 = 74.7250$ .

We also present the expression for computing Baseline corrected Tau, given that it is not available in Tarlow (2017). The index is computed, either on the corrected data  $\tilde{y}$  or on the original data  $y_t$ , as Kendall’s (1970) tau correlation coefficient (specifically,  $\tau_B$ ) between the measurements and a dummy variable Phase with  $n_A$  values equal to 0 followed by  $n_B$  values equal to 1. Formally, the result can be obtained as:

$$\frac{[(y_g > y_h) - (y_g < y_h)]}{\sqrt{[\frac{n(n-1)}{2} - T_X][\frac{n(n-1)}{2} - T_Y]}} \quad (23)$$

where  $n = n_A + n_B$ ,  $T_X = \frac{n_A(n_A-1)}{2} + \frac{n_B(n_B-1)}{2}$ , representing the ties in the *Phase* variable, and  $T_Y = \sum_{i=1}^m \frac{f_i(f_i-1)}{2}$ , representing the ties in the outcome variable Y, with  $f_i$  being the frequency of each of the  $m$  different values of Y. For instance, if the Y values are 8, 5, 7, 7, 8, and 7, there are three different values (5, 7, and 8) and thus  $m = 3$ , with  $f_1 = 1$  (5 occurs once),  $f_2 = 3$  (7 occurs three times), and  $f_3 = 2$  (8 occurs twice).

For the running example we used the transformed data (71.4075, 81.5850, 115.0525, 120.0700, 112.1475, 70.8450, 72.8225, 82.2600, 45.1575, 55.6150, 59.0825, 71.5200, 76.4775, and 74.7250). There are  $(n(n-1))/2 = (14(14-1))/2 = 91$  comparisons performed. From these comparisons, there are 8 cases in which a phase B corrected value is greater than a phase A corrected value and 40 cases in which a phase B corrected value is smaller than a phase A corrected value. The number of ties in the Phase variable is  $T_X = \frac{n_A(n_A-1)}{2} + \frac{n_B(n_B-1)}{2} = \frac{8(8-1)}{2} + \frac{6(6-1)}{2} = 28 + 15 = 43$ . The number of ties in the outcome variable Y is 0, because all the values appear only once ( $m = 14$  and all  $f_i = 1$ ). Thus, Baseline corrected Tau is equal to:

$$\frac{[(y_g > y_h) - (y_g < y_h)]}{\sqrt{[\frac{n(n-1)}{2} - T_X][\frac{n(n-1)}{2} - T_Y]}} = \frac{8 - 40}{\sqrt{(91 - 43)(91 - 0)}} = \frac{-32}{66.06} \simeq -0.484.$$



## 4.7 Detrending in the Between-Case Standardized Mean Difference (BC-SMD)

The BC-SMD is mainly applicable to stable data, but it is possible to detrend the measurements before obtaining its value (Shadish, Hedges, & Pustejovsky, 2014). Shadish et al. (2014) and Marso and Shadish (2015) comment that one possible way of detrending is to first run the model

$$y_t = b_0 + b_1t + b_2Phase \times t \quad (24)$$

and then to perform the main analysis on the residuals of the previous analysis:  $e_t = y_t - \hat{y}_t$ . Additionally the researcher may center the measurement occasion variable  $t$ , according to substantive criteria (Marso & Shadish, 2015), for instance,  $t.cent = t - (n_A + 1)$  would lead to the value of 0 being assigned to the first intervention phase measurement occasion.

For the running example, the regression model from Equation (24) leads to  $b_0 = 88.85$ ,  $b_1 = -4.94$ , and  $b_2 = -2.07$ . The fitted values are, therefore,  $y_1 = 88.85 + (-4.94) \times 1 + (-2.07) \times 0 \times 1 = 83.91$ ,  $y_2 = 88.85 + (-4.94) \times 2 + (-2.07) \times 0 \times 2 = 78.96$ , up to  $y_{14} = 88.85 + (-4.94) \times 14 + (-2.07) \times 1 \times 14 = -9.35$ .  $e_1 = y_1 - \hat{y}_1 = 66.39 - 83.91 = -17.52$ ,  $e_2 = y_2 - \hat{y}_2 = 71.55 - 78.96 = -7.441$ , up to  $e_{14} = y_{14} - \hat{y}_{14} = 4.48 - (-9.35) = 13.83$ . The main analysis is performed with the residuals.

## 4.8 Slope and Level Change (SLC)

In SLC, baseline trend estimated according to Equation (9) is removed from the baseline and intervention phase measurements using the expression:

$$\tilde{y}_t = y_t - b_{1(D)}t; t = 1, \dots, n_A + n_B \quad (25)$$

This expression is equivalent to Equation (22) but for the estimate of slope used. Thus, just like for the Baseline corrected Tau, the original measurement units of the dependent variable are maintained.

Using the detrended data, the following steps in SLC are: (a) to estimate the slope of the trend line fitted to the intervention using Equation (9), which represents the change in slope (SC); (b) remove the trend from the intervention phase via  $\tilde{\tilde{y}}_g = \tilde{y}_g - (i - 1) \times SC$ ;  $g = 1, 2, \dots, n_B$ ; (c) estimate the net change in level as the difference between the average of double-detrended intervention phase data  $\tilde{\tilde{y}}_g$  and the average of the detrended baseline data  $\tilde{y}_h$  (where  $h = 1, 2, \dots, n_A$ ).

For the running example, we already obtained  $b_{1(D)} = -3.47$  from the baseline data. The detrended baseline and intervention phase data is, therefore, obtained as  $\tilde{y}_1 = y_1 - b_{1(D)} \times 1 = 66.39 - (-3.47) \times 1 = 69.86$ ,  $\tilde{y}_2 = y_2 - b_{1(D)} \times 2 = 71.55 - (-3.47) \times 2 = 78.49$ , up to  $\tilde{y}_{14} = y_{14} - b_{1(D)} \times 14 = 4.48 - (-3.47) \times 14 = 53.06$ . The subsequent analyses are performed with the detrended data.

#### 4.8.1 Summary table of results after detrending

Table C.1 illustrates the different results to which a quantification such as the mean difference would lead to, after following the different approaches for controlling for baseline trend.

**Table C.1.** Baseline and intervention means of the Figure 1, detrending according to several different approaches.

Detrending approach	Baseline phase (A) mean	Intervention phase (B) mean	Mean difference (B-A)
None: Original data	68.195	6.061	-62.134
Allison and Gorman	0.00	-17.978	-17.978
Mean and Slope Adjustment	68.195	50.217	-17.978
Trend analysis by Gorsuch	3.782	-5.043	-8.825
Differencing analysis by Gorsuch	-3467	-6.273	-2.806
Slope and level change	83.797	45.934	-37.863
Detrending from BC-SMD	1.591	-2.121	-3.731
Baseline corrected Tau	90.774	63.763	-27.010

Gorsuch's trend analysis applied by removing overall trend

## 5 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.
- Bartlett, M. S. (1949). Fitting a straight line when both variables are subject to error. *Biometrics*, 5, 207-212.
- Brossart, D. F., Vannest, K., Davis, J., Patience, M. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation*, 24, 464-491.
- Bulté, I., Onghena, P. (2012). When the truth hits you between the eyes: A software tool for the visual analysis of single-case experimental data. *Methodology*, 8, 104-114.
- Chatfield, C. (2000). *Time-series forecasting*. London, UK: Chapman Hall/CRC.
- Eilers, H. J., & Hayes, S. C. (2015). Exposure and response prevention therapy with cognitive defusion exercises to reduce repetitive and restrictive behaviors displayed by children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 19, 18-31.
- Gast, D. L., & Spriggs, A. D. (2010). Visual analysis of graphic data. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 199-233). London, UK: Routledge.
- Gibson, W. M., & Jowett, G. H. (1957). Three-group regression analysis, Part I. Simple regression analysis. *Applied Statistics*, 6, 114-122.
- Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment*, 5, 141-154.
- 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.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
- Johnstone, I. M., & Velleman, P. F. (1985). The resistant line and related regression methods. *Journal of the American Statistical Association*, 80, 1041-1054.
- Kendall, M. G. (1970). *Rank correlation methods*. London, UK: Griffin.
- Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation*, 24, 445-463.
- Manolov, R., Rochat, L. (2015). Further developments in summarising and meta-analysing single-case data: An illustration with neurobehavioural interventions in acquired brain injury. *Neuropsychological Rehabilitation*, 25, 637-662.
- Manolov, R., & Solanas, A. (2013). A comparison of mean phase difference and generalized least squares for analyzing single-case data. *Journal of School Psychology*, 51, 201-215.
- Marso, D., Shadish, W. R. (2015). *Software for meta-analysis of single-case design: DHPS macro*. Retrieved January 22, 2017 from <http://faculty.ucmerced.edu/wshadish/software/software-meta-analysis-single-case-design>
- McCabe, G. P., Jr. (1980). The use of the 27% rule in experimental design. *Communications in Statistics, Part A: Theory and Methods*, 9, 765-776.
- Mendenhall, W., & Sincich, T. (2012). *A second course in statistics: Regression analysis* (7th ed.). Boston, MA: Prentice Hall.
- Miller, M. J. (1985). Analyzing client change graphically. *Journal of Counseling and Development*, 63, 491-494.
- 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., Cryer, J., & Byrns, G. (2006). Controlling baseline trend in single-case research. *School Psychology Quarterly*, 21, 418-443.
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2014). A simple method to control positive baseline trend within data nonoverlap. *The Journal of Special Education*, 48, 79-91.
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, 42, 284-299.
- Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *American Statistical Association Journal*, 63, 1379-1389.
- Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology*, 52, 123-147.

- Solanas, A., Manolov, R., & Onghena, P. (2010). Estimating slope and level change in N=1 designs. *Behavior Modification*, 34, 195-218.
- Tarlow, K. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected Tau. *Behavior Modification*, 41, 427-467.
- Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis, III. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen [Proceedings of the Royal Netherlands Academy of Sciences]*, 53, 1397-1412. Retrieved on January 9, 2017 from <http://oai.cwi.nl/oai/asset/8269/8269A.pdf>
- Vannest, K. J., Davis, J. L., & Parker, R. I. (2013). *Single case research case in schools: Practical guidelines for school-based professionals*. New York, NY: Routledge.
- Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2016). *Single Case Research: web based calculators for SCR analysis*. (Version 2.0) [Web-based application]. College Station, TX: Texas AM University. Available from [singlecaseresearch.org](http://singlecaseresearch.org)
- Velleman, P. F., & Hoaglin, D. C. (1981). *Applications, basics and computing of exploratory data analysis*. Boston, MA: Duxbury Press.
- White, O. R., & Haring, N. G. (1980). *Exceptional teaching: A multimedia training package*. Columbus, OH: Merrill.
- Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). London, UK: Academic Press.