Turkish Journal of Psychology, June 2020, *35*(85), 61-64 DOI: 10.31828/tpd1300443320181023m000015

Summary
Effect Size Calculations in Single-Case Experimental Research Design: An Investigation of Nonoverlap Methods

Salih Rakap
Ondokuz Mayıs University

Şerife Yücesoy-Özkan
Anadolu University

Sinan Kalkan
Çanakkale Onsekiz Mart University

Initially used in the medical field in the 1970s, evidence-based practices have become important in many areas including psychology, health sciences, nursing, and educational sciences (Odom et al. 2005; Rakap, 2016; Reichow, 2016). No Child Left Behind (NCLB, 2002) law amended in 2002 brought radical changes in education in the United States. NCLB required all schools and teachers to use evidence-based practices while teaching. Following the passage of this law, many researchers and organizations in the United States and around the world began to work on determining evidence-based practices in general and special education through rigorous meta-analysis of previous research (Odom et al., 2005; Olive & Franco, 2007).

An important limitation of these meta-analyses has been the exclusion of studies employing single-case research (SCER) designs (Maggin, Briesch, & Chafouleas, 2013; Rakap, Snyder, & Pasia, 2014; Wolery, Busick, Reichow, & Barton, 2010). SCER designs are commonly used in psychology, social work, special education, and applied behavior analysis disciplines and allow rigorous investigation of the causal or functional relationship between dependent and independent variables. The main reason for the exclusion of SCER studies from meta-analyses of evidence-based practices is the lack of agreed-upon methods to calculate effect size estimates for data obtained from SCER studies (Campbell, 2004; Parker & Hagan-Burke, 2007; Rakap et al., 2014; Wolery et al., 2010). To overcome this problem, several research groups have been working on developing effect size calculation methods for SCER. These methods can be grouped under three main categories: parametric methods (Center, Skiba, & Casey, 1985; Maggin et al., 2013; Van den Noortgate & Onghena, 2003), standardized mean difference (Hedges, Pustejovsky, & Shadish, 2012), and nonoverlap methods (Parker et al., 2011; Scruggs et al., 1987).

International literature includes a number of different nonoverlap methods that are similar but have small differences. These include Percentage of Nonoverlapping Data (PND; Scruggs et al., 1987), Percentage of Zero Data (PZD; Scotti, Evans, Meyer, & Walker, 1991), Percentage of Data Exceeding the Median (PEM; Ma, 2006), Percentage of All Nonoverlapping Data (PAND; Parker & Hagan-Burke, 2007), Nonoverlap of All Pairs (NAP; Parker & Vannest, 2009), Improvement Rate Difference (IRD; Parker, Vannest, & Brown, 2009), Percentage of Data Exceeding a Median Trend (PEM-T; White & Haring, 1980; Wolery et al., 2010), Pairwise Data Overlap (PDO; Parker & Vannest, 2007), Pairwise Data Overlap Squared (PDO^2; Wolery et al., 2010), Tau_{Novlap} (Parker, Vannest, Davis et al., 2011), Tau-*U* (Parker, Vannest, Davis et al., 2011) and Phi (Parker, Vannest, & Davis, 2011). Although methods developed more recently are more sophisticated, many national and international researchers still use early methods, such as PND or PZD when calculating effect size estimates in SCER (e.g., Bozkus-Genc & Yucesoy-Ozkan, 2016; Sonmez & Diken, 2010; Sazak-Pinar & Merdan, 2016; Tavil & Karasu, 2013; Toper-Korkmaz & Diken, 2010; Yucesoy-Ozkan & Sonmez, 2011).

Considering the emphasize of the American Psychology Association (APA) and What Works Clearinghouse (WWC) on the use of effect size estimates to supplement the findings of studies using group or SCER, it has become important for national researchers to know characteristics of nonoverlap methods, their advantages and limitations, and when to use which methods based on data patterns. The purpose of the present study is to describe 13 nonoverlap methods used to calculate effect size estimates in SCER, explain their calculations using hypothetical data sets, discuss their advantages and disadvantages, and provide a guideline for interpreting effect size estimates and considerations for selecting appropriate effect size methods based data obtained from SCER studies.

Address for Correspondence: Asst. Prof. Salih Rakap, Ondokuz Mayıs University, Faculty of Education, Special Education, Atakum / Samsun
E-mail: salih.rakap@omu.edu.tr

Effect Size Methods and Calculations

Percentage of Nonoverlapping Data

Percentage of nonoverlapping data (PND; Scruggs et al., 1987) is the first method developed to calculate effect size estimates in SCER. PND is conceptualized as the percentage of data points in the intervention phase that have higher (or lower for studies aiming to reduce behavior) numeric value than the highest datum point in the baseline phase (Parker, Vannest, & Davis, 2011; Scruggs et al., 1987; Scruggs & Mastropieri, 2013). PND effect size estimate is calculated using the following steps: (a) determining the datum point with the highest numeric value in the baseline phase, (b) drawing a line (nonoverlap line) that is parallel to x-axis from this datum point towards the intervention phase, (c) counting the number of data points in the intervention phase that remain above the nonoverlap line, (d) dividing the sum obtain in step c to total number of data points in the intervention phase, and (e) multiplying the result obtained in step d by 100 (Scruggs & Mastropieri, 1998; Scruggs et al., 1987).

Percentage of Data Exceeding the Median

Percentage of data exceeding the median (PEM; Ma, 2006) is conceptualized as the percentage of data points in the intervention phase that is above the median line drawn based on the baseline phase data. PEM effect size estimate is calculated by (a) determining the median of the baseline phase data, (b) drawing a line (median line) that is parallel to x-axis from this datum point towards the intervention phase, (c) counting the number of data points in the intervention phase that remain above the median line, (d) dividing the sum obtain in step c to total number of data points in the intervention phase (Ma, 2006), and (e) multiplying the result obtained in step d by 100 (Parker, Vannest, & Davis, 2011).

Percentage of Data Exceeding a Median Trend

Percentage of data exceeding a median trend (PEM-T; White & Haring, 1980; Wolery et al., 2010) is conceptualized as the percentage of data in the intervention phase that is above the trend line drawn based on data in the baseline phase. PEM-T effect size estimate is calculated by (a) drawing a line from the middle point of the first two data points to the last two data points in the baseline phase (trend line) and extending the line across the intervention phase, (b) counting the number of data points in the intervention phase that remain above the trend line, (c) dividing the sum obtain in step b to total number of data points in the intervention phase, and (d) multiplying the result obtained in step c by 100 (Rakap et al., 2014; Wolery et al., 2010).

Percentage of Zero Data

Percentage of zero data (PZD; (Scotti et al., 1991) is the only nonoverlap method specifically developed for calculating effect size estimates for SSER studies aiming to decrease behaviors. PZD is calculated by finding the first data point in the intervention phase that reaches zero and calculating the percentage of data points obtained in the intervention phase, including the first zero point, that remains at zero (Scotti et al., 1991).

Percentage of All Nonoverlapping Data

Percentage of all nonoverlapping data (PAND; Parker & Hagan-Burke, 2007) is conceptualized as the percentage of data remaining in the baseline and intervention phases after the minimum number of data points is removed to eliminate the overlap between the baseline and intervention phases. PAND is calculated by (a) determining the total number of data points across the baseline and intervention phases, (b) determining minimum number of data points that eliminates the overlap between the baseline and intervention phases, (c) subtracting the number obtained in step b from the sum obtained in step a, (d) dividing the result obtained in step c by the sum obtained in step a, and (e) multiplied the quotient obtained in step d by 100 (Parker & Hagan-Burke, 2007; Parker, Vannest, & Davis, 2011).

Percentage of Non-overlapping Corrected Data

Percentage of non-overlapping corrected data (PNCD; Manalov & Solonas, 2009) is developed to overcome the limitations of PND by offering a correction procedure implemented prior to the application of PDN. PNCD aims to control for unwanted trend observed in baseline phase. PNCD is calculated by (a) subtracting each datum point in baseline phase from the consecutive data point to create a difference series with n_A-1 data points, (b) calculating mean of the difference series, (c) computing the trend-correction factor for each datum point across phases by multiplying the mean calculated in step b with the sequence number of each datum point, (d) subtracting the trend-correction factor calculated for each datum point from the original value of the datum point, and (e) applying the PND procedure to calculate PNCD.

Improvement Rate Difference

Improvement rate difference (IRD; Parker, Vannest, & Brown, 2009) was initially developed in the medical field as risk-reduction or risk difference method. IRD is conceptualized as the difference between the improvement rates of the baseline and intervention phases. To calculate IRD, first, the total number of data points across baseline and intervention phases is determined,

and then the minimum number of data points that eliminate the overlap between the baseline and intervention phases is determined. When all data points are removed from the intervention phase to eliminate the overlap, the IRD is calculated by (a) dividing the number of remaining data points in the intervention phase after removal to the total number of data points in the intervention phase, and (b) multiplying the quotient by 100. When all data points are removed from the baseline phase, the IRD is calculated by (a) dividing the number of remaining data points in the baseline phase after removal to the total number of data points in the baseline phase, (b) subtracting the quotient from 1, and (c) multiplying the result by 100. When the data points are removed both from baseline and intervention phases to eliminate the overlap, IRD s calculated by (a) dividing the number of remaining data points in the intervention phase by the total number of data points in the intervention phase, (b) dividing the number of data points removed from the baseline phase by the total number of data points in the baseline phase, (c) subtracting the quotient from step a from step b, and (d) multiplying the result by 100 (Parker et al., 2009; Vannest & Ninci, 2015).

Pearson's Phi

Pearson's Phi (Parker, Vannest, & Davis, 2011) is developed as an extension to PAND. Phi is calculated by (a) determining the minimum number of data points removed from the baseline and intervention phases to eliminate the overlap between phases, (b) using half of the number of data points removed to create ratios for the baseline and intervention phases, (c) submitting these two ratios in 2 × 2 table to cross-tab analysis, and (d) multiplying the result by 100 (Parker, Vannest, & Davis, 2011).

Nonoverlap of All Pairs

Nonoverlap of all pairs (NAP; Parker & Vannest, 2009) is conceptualized as the percentage of all pairwise comparisons across the baseline and intervention phases that show improvement from the baseline phase to the intervention phase. NAP is calculated in four steps: (a) determining the total number of pairs by multiplying the number of data points in the baseline and intervention phases, (b) comparing each datum point in the baseline phase with each datum point in the intervention phase to determine the number of pairs in which the intervention phase datum point has a higher numeric value than the datum point in the baseline phase (improving data points) and the number of pairs in which numeric values of the data points across phases are equal (tie), (c) summing the total number of improved pairs and half of the pairs with tie, (d) dividing the sum obtained in step c by the total number of pairs, and (e) multiplying the quotient by 100 (Parker &Vannest, 2009).

Pairwise Data Overlap and Pairwise Data Overlap Squared

Pairwise data overlap (PDO; Parker & Vannest, 2007) is also conceptualized as the percentage of all pairwise comparisons across the baseline and intervention phases that show improvement from the baseline phase to the intervention phase. Calculation of effect size estimates using PDO is very similar to the calculation of effect size estimates using NAP; the only difference between the two methods is that PDO does not take ties (pairs with equal values) into account. PDO^2 (Wolery et al., 2010) is calculated by squaring the effect size estimate obtained from PDO analysis.

Tau_{Novlap}

Tau_{Novlap} (Parker, Vannest, Davis et al., 2011) is also based on pairwise comparisons of data as in NAP and PDO. To calculate effect size estimates using Tau_{Novlap}, each datum point in the baseline phase is compared to each datum point in the intervention phase. Next, the total number of decreasing pairs (i.e., when a baseline datum point > an intervention datum point) is subtracted from the total number of pairs with improvement (i.e., when an intervention datum point > a baseline datum point). The result is divided by the total number of pairs, and resulting quotient is multiplied by 100 (Parker, Vannest, Davis et al., 2011).

Tau-U

Tau-*U* (Parker, Vannest, Davis et al., 2011) is the second nonoverlap method that can control for a baseline trend in a therapeutic direction. Tau-*U* is the most sophisticated nonoverlap method to calculate effect size estimates in SCER. Calculation of Tau-*U* statistics is more complex than the calculations of effect size estimates using the other nonoverlap methods. Tau-*U* can be considered as an extension to Kendall's rank order correlation and Mann-Whitney U test. To calculate effect size estimates using Tau-*U*, (a) the total number of pairs is determined by multiplying the number of data points in baseline and intervention phases, (b) Kendall's rank order correlation analysis is conducted to obtain S value, (c) S value is divided by the total number of pairs calculated in step a, and (d) the quotient is multiplied by 100 (Parker, Vannest, Davis et al., 2011).

Interpreting Effect Size Estimates

Researchers have developed benchmarks for interpreting effect size estimates calculated using different

methods. Effect size calculation methods such as PND, PZD, IRD, NAP, and Tau-*U* have their own benchmarks specifically developed for these methods. However, methods such as PEM, PEM-T, PAND, PDO, PDO^2, and Tau_{Novlap} do not have their own criteria to evaluate obtained effect sizes, and many researchers use benchmarks developed for PND while interpreting effect size estimates calculated using these methods. This may be problematic especially when the characteristics of effect size calculation methods are considerably different from PND. Researchers must be cautious when using benchmarks developed for another method in their research. Moreover, effect size estimates should never be used alone to evaluate the effectiveness of an intervention in SSER studies; they should always be used in conjunction with visual analysis.

Effect Size Estimates and Visual Analysis

Numerous studies have been conducted to investigate the relationships between the magnitude of effect size estimates and results of visual analyses since the inception of the PND (e.g., Ma, 2006; Parker & Hagan-Burke, 2007; Parker & Vannest, 2009; Rakap et al., 2014; Rakap et al., 2017; Wolery et al., 2010). For example, Ma (2006) reported that the correlation between PEM and visual analysis results were higher than the correlation between PND and visual analysis. Parker and Hagan-Burke (2007) found that IRD had the highest correlation with visual analysis while PEM had the lowest correlation with the results of the visual analysis. In two different studies, Rakap and colleagues (2014; 2017) compared the effect size estimates calculated by 13 different nonoverlap methods and visual analysis results using the same data set and reported that PND was the most compatible effect size methods with visual analysis followed by PEM-T, NAP, Tau_{NOVLAP}, IRD, and PAND, while PDO^2 was the least compatible method. More research is needed to determine the most compatible effect size methods with visual analysis using different datasets.

Considerations for Selecting Effect Size Methods

When selecting the appropriate effect size method for a SCER study, several characteristics of data and data patterns (e.g., number of data points in each phase, a baseline trend in therapeutic direction, and overlap level) should be considered (Brossart et al., 2014). A number of data points in each phase significantly influence the effect size estimates calculated using nonoverlap methods. Although three or five data points in the baseline phase are acceptable in SCER, a stable data patterns should be obtained before an intervention is implemented (Brossart et al., 2014; Lenz, 2013). Moreover, when a baseline trend in therapeutic direction exists, baseline phase should be extended until a stable pattern of data is observed and an effect size estimate (e.g., Tau-*U* or PEM-T) that controls for the baseline trend should be used to calculate effect sizes (Brossart et al., 2014). In data sets where there is a lot of overlapping data between the baseline and intervention phases, one of the effect size methods that allows comparison of individual data points across phases should be used. Overall, considering their consistency with the results of visual analysis and ability to control for some of the characteristics of data obtained from SCER studies, it can be recommended that PEM-T or Tau-*U*, along with PND, NAP or IRD should be used together to calculate and report effect sizes estimates in SCER.

Conclusions

The need for using effect size estimates to determine the experimental effect in SCER is increasing within the context of identifying evidence-based practices; however, there is no consensus about which effect size methods should be used to calculate effect sizes in SCER. In this respect, there is a need for more studies comparing and contrasting parametric methods, methods based standardized mean difference, and nonoverlap methods as well as their relationships with visual analysis. Until a method that controls for all characteristics and data patterns (e.g., level, trend and variability) of SCER is developed, more than one method to calculate and report effect size estimates should be used (Brossart et al., 2014; Kratochwill et a., 2010; Rakap et al., 2014). Nevertheless, an important point that should always be remembered under any circumstances is that the primary method for determining causal or functional relationships between dependent and independent variables in SCER is visual analysis and effect size estimates must be used to supplement and support findings of the visual analysis. Moreover, when interpreting the results of visual analysis and effect size calculations, researchers should consider the context in which the study is conducted and clinical importance of the target behavior to be changed.