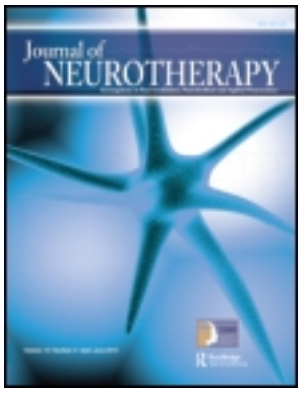


Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wneu20

Single-Case Design in Psychophysiological Research: Part II: Statistical Analytic Approaches

Dwight E. Waddell [a], Stephanie L. Nassar [a] & Scott A. Gustafson [a]

[a] The University of Mississippi, University, Mississippi, USA

Available online: 20 May 2011

To cite this article: Dwight E. Waddell, Stephanie L. Nassar & Scott A. Gustafson (2011): Single-Case Design in Psychophysiological Research: Part II: Statistical Analytic Approaches, Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience, 15:2, 160-169

To link to this article: http://dx.doi.org/10.1080/10874208.2011.570693

Journal of Neurotherapy, 15:160–169, 2011

ISSN: 1087-4208 print/1530-017X online
DOI: 10.1080/10874208.2011.570693

Routledge
Taylor & Francis Group

SINGLE-CASE DESIGN IN PSYCHOPHYSIOLOGICAL RESEARCH: PART II: STATISTICAL ANALYTIC APPROACHES

Dwight E. Waddell, Stephanie L. Nassar, Scott A. Gustafson

The University of Mississippi, University, Mississippi, USA

This article is Part 2 of a two-part series on the conceptual and methodological applications of single-case design research in psychophysiological research (Gustafson, Nassar, & Waddell, 2011). Part 1 in this series presented the context, structure, and fundamentals of single-case design in relation to psychophysiology. Part 2 introduces four statistical analyses that are utilized in single-case research design and are broadly applicable to a wide range of research questions or clinical outcome studies. These techniques are reviewed in sufficient detail so that clinicians and researchers may apply them in real-world contexts. The following analyses—(a) Percentage of Non-overlapping Data Points and Percentage of All Non-overlapping Data, (b) Split-Middle and Percentage of Data Points Exceeding the Median, (c) Improvement Rate Difference, and (d) Hierarchical Linear Modeling—were chosen for their suitability with psychophysiological data. Although these analyses may be unfamiliar, their calculations are quite straightforward. Special emphasis is given to statistics that provide effect size data, as this statistic allows studies to be incorporated in to meta-analytic studies, promoting cumulative knowledge across time.

In Part 1 of this two-part series, the conceptual and methodological applications of single-case design research methodology were examined in detail (Gustafson, Nassar, & Waddell, 2011). The current psychophysiological research paradigm of large-scale, group comparison designs may not be the best conceptual fit for studying psychophysiological phenomena. Clinical psychophysiology employs the reinforcement of successive, incremental changes of specific behaviors, such as producing greater amplitude of a particular electroencephalographic (EEG) bandwidth. The nature of this intervention is closer to operant conditioning and physical rehabilitation than the commonly used standard pharmacological model, which employs large total sample size, group comparison designs.

To review briefly, single-case design allows researchers and clinicians to make reasonable determinations about the efficacy and the effectiveness of particular interventions using a small number of experimental subjects. The foundation of this design is determining cause and effect by obtaining a preintervention baseline, systematic sampling during active intervention (such as neurofeedback), and implementing either a reversal or recording without feedback phase (Figure 1).

The resulting data stream contains many useful features, such as the angle of change from one phase to the other, the magnitude of change, the number of sessions required for observable changes to take place, and the effects of withdrawing feedback or instituting bidirectional training. Visual inspection of the data is often clinically informative, but the gathering of data points that are connected to each other in time and in sequence presents a set of analytical challenges that may be unfamiliar to some clinician-researchers.

Received 25 January 2011; accepted 28 February 2011.
Address correspondence to Scott A. Gustafson, PhD, Psychological Services Center, Kinard Hall, Suite G-382, University of Mississippi, University, MS 38677, USA. E-mail: sagustaf@olemiss.edu

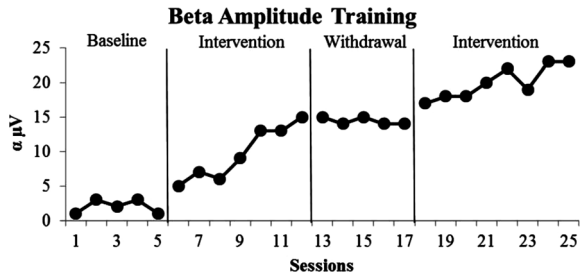


FIGURE 1. Data from a standard single-case A/B/A/B design. *Note.* Baseline indicates no feedback. Intervention indicates operantly reinforcing beta amplitude, withdrawal indicates recording without feedback and intervention demonstrates a continuation of the intervention.

TIME SERIES DATA AND NONPARAMETRIC ANALYSES

One of the primary issues that can act as a barrier to clinician-scientists utilizing single-case design is the problem of how to statistically analyze the data. It is perhaps the most confusing, daunting, and disjointed element of this experimental method. As a matter of fact, until very recently there has been an active debate whether statistical analyses are even necessary for this kind of data, with some arguing for simple visual inspection of the data rather than performing mathematical analyses (Kazdin, 1982; Parker, Hagan-Burke, & Vannest, 2007). There is a growing consensus, however, that the inclusion of quantitative analyses holds many advantages, including increased reliability of results reporting, reduction of Type I and Type II errors, greater discrimination of treatment effects, and greater confidence in conclusions about cause and effect (Franklin, Gorman, Beasley, & Allison, 1996; Parker, Hagan-Burke, & Vannest, 2007).

Even when one acknowledges the need for empirical analyses, there is a confusing array of potential analytical approaches, formulae, and debate about the appropriateness of certain statistics in certain situations (Jones, 2003; Parker & Brossart, 2003). This is largely due to the unique nature of the data. Single-case design entails sequential observations over time, generally to determine incremental changes in behavior as a result of a specific intervention (Gustafson et al., 2011).

It is by nature a time-series data set, which means that the data points will be serially dependent. In other words, each data point in the series is correlated with the next data point, allowing for an amount of prediction from one measurement epoch to another. The extent of this predictability can be determined by examining the data's autocorrelation: the degree to which the data correlates with itself over sequential time intervals (Kazdin, 1984).

Serially dependent data make statistical analysis somewhat unusual in that they violate the *Independence of Error* assumption. This is one of the few violations of assumptions to which parametric statistics, such as analysis of variance (ANOVA) and correlation, are not robust (Borckardt & Nash, 2002). To better understand the context for this limitation, remember that statistics based on the General Linear Model (such as ANOVA, correlations, etc.) require random or independent samples from a population. This generally results in observations that fall along a normally distributed curve. Autocorrelated data, being taken serially, share variance and error from one sample to the next. The resulting data will not be normally distributed. Gaussian distributions are a fundamental underlying assumption of the General Linear Model. This effectively rules out most *parametric* statistics, which are the exact analytical techniques most clinicians and researchers are trained to conduct.

Fortunately, there has been a great increase in the availability, ease of use, and sophistication of *nonparametric* analyses within the past 10 years that are directly applicable to single-case design data sets (Jenson, Clark, Kircher, & Kristjansson, 2007). These nonparametric statistical analyses vastly increase the researcher's ability to examine experimental significance, clinical significance, and effect size of a clinical intervention. The inclusion of effect size indices, contained in most recent analytical approaches, facilitates the ability of independent researchers to conduct meta-analytical techniques, which compounds an individual study's contribution to the scientific literature (American Psychological Association [APA], 2001; Campbell, 2004).

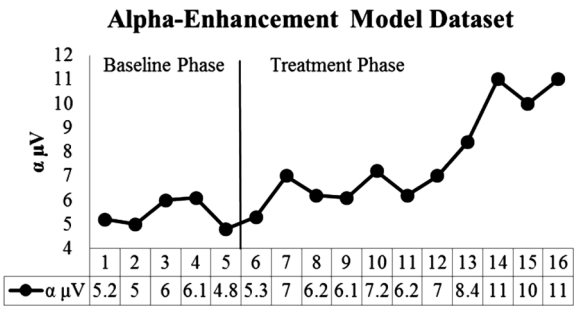
FOUR STATISTICAL TECHNIQUES

There is a wide variety of nonparametric statistics from the very basic to the very sophisticated (Hays, 1994). Four statistical techniques are presented here, chosen for their direct applicability to psychophysiological research and appropriateness for a broad range of experimental designs. Although the following statistical techniques have rather simple calculations, they have been demonstrated to be better suited to single-case design data sets than a number of more sophisticated techniques. They are more robust to serial dependency effects, are more sensitive to changes in effect size (Manolov & Solanas, 2008), and have greater statistical power compared to parametric analyses (Jenson et al., 2007).

Perhaps their greatest advantage over other potential analytical approaches is that three of the four techniques presented here allow for calculation of the *effect size* statistic. This statistic is a clinical tool to estimate the relative magnitude of treatment effects across studies (e.g., the degree to which an intervention exerts change on a target behavior). The effect size statistic also makes the results of the study eligible for inclusion in meta-analytic research, which is a critical component of empirically demonstrating fidelity of treatment approaches across time and situations (APA, 2001; Campbell, 2004).

These approaches have been endorsed by recognized experts in scientific design that range the methodological spectrum from visual inspection without quantitative analysis (Kazdin, 1982) to meta-analysis (Kavale, Mathur, Forness, Quinn, & Rutherford, 2000). In addition, the calculations are so straightforward, they usually do not require specialized statistical software.

These techniques are uniquely well suited to psychophysiological research and are relatively straightforward to understand and employ. The following review presents a selected cross-section of four of the most useful and accessible techniques: Percentage of Non-overlapping Data Points (PND) and Percentage of All Non-Overlapping Data (PAND),



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
α μV	5.2	5	6	6.1	4.8	5.3	7	6.2	6.1	7.2	6.2	7	8.4	11	10	11

FIGURE 2. Alpha-enhancement Model Dataset. *Note.* All the statistical techniques presented in this paper utilize this model dataset for conceptual clarity and comparison among techniques. The data presented are of an actual neurofeedback training protocol conducted for the purposes of this article.

Split-Middle and Percentage of Data Points Exceeding the Median (PEM) Methods, Improvement Rate Difference (IRD) analyses, and Hierarchical Linear Modeling (HLM). These empirical techniques are presented in enough detail (i.e., step-by-step instructions and demonstration graphs) as to allow the clinician-researcher to conduct these analyses with his or her own unique single-case design data sets. To enhance the readability and comparison of these four statistical techniques, the authors have elected to employ the same electroencephalographic dataset (Figure 2) for each calculation.

PND AND PAND

PND (Scruggs, Mastropieri, & Casto, 1987) and its variant, PAND (Parker et al., 2007), have emerged as two of the most common statistical approaches to single-case design (Scruggs & Mastropieri, 2001). Both are straightforward, are simple to calculate, and provide meaningful results.

PND and PAND are predicated on the assumption that the difference between the baseline phase and intervention phase of a single-case design will be of sufficient magnitude to ensure that data points will not significantly overlap. The lowest point during a treatment phase is expected to be generally higher than the highest point of the baseline phase (Figure 3).

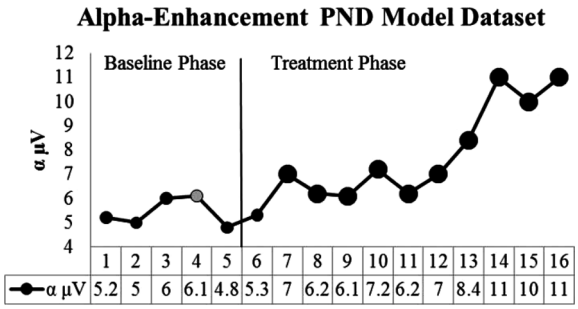


FIGURE 3. Percentage of Non-Overlapping Data Points (PND) calculation method utilizing the model data set. *Note*. This approach measures the number of observations in the treatment phase that exceed the highest point in the baseline phase (gray marker). Data points counted in the treatment condition are noted by larger markers ($n = 10$).

The calculation method for PND is simple:

1. Calculate the number of Intervention Phase observations, or data points, that are of greater magnitude than the highest data point in the Baseline Phase:
   Example results: 10
2. Divide this value by the number of observations in the Intervention Phase (resulting in a proportion):
   Example results: $10 \div 11 = .9091$
3. Multiply the proportion by 100 to obtain a percentage:
   Example results: $.9091 \times 100 = 90.91\%$
4. Scruggs and Mastropieri (1994) provided general interpretational guidelines for PND results:
   - $PND > 70$ for effective treatments
   - $70 > PND > 50$ for questionable effectiveness
   - $PND < 50$ for no effective results in the current example, $PND > 70$, resulting in an interpretation of ''effective.''

The trade-off for this elegance and simplicity is a somewhat limited utility. Shortcomings of PND include its lack of a common effect size statistic, which makes inclusion in meta-analytic studies and direct comparisons to other interventions difficult (Strain, Kohler, & Gresham, 1998). It is also of limited utility in more complicated designs other than a simple AB design. For subsequent baseline or intervention phases, the choices are to compare later trials to the original baseline or calculate multiple results for every baseline/intervention pair, which may skew the results (Jenson et al., 2007).

Last, and of particular concern to clinician-researchers in psychophysiology, care must be taken when choosing appropriate data sets. PND is vulnerable to false negatives in data with high degrees of variability. This may be more acceptable when examining data that do not ''bounce'' much between time frames, such as extremity temperature, but may cause significant problems when examining high variance data sets, such as nonsmoothed neurofeedback of EEG spectral frequencies.

However, minor calculation adjustments can produce dramatic increases in the utility of results. Given its simplicity, elegance, and interpretability, PND's advantages seem to outweigh its shortcomings, as evidenced by its use in the published literature.

An alternative to PND that addresses many of its limitations is PAND. PAND can be applied to multiple baseline designs, is useful for longer data sets, and allows for the calculation of Pearson's phi, a useful effect size statistic. The primary difference is that instead of using one data point in the Baseline Phase, all available data are utilized. A step-by-step set of instructions for calculating PAND in a spreadsheet format is listed in the appendix of Parker, Vannest, and Brown (2009).

PEM

White and Haring's (1980) Split-Middle Method and Ma's (2006) PEM approach data slightly differently, using the *median* of Phase A (Baseline) scores as the basis of comparison, as opposed to the highest data point. This simple adjustment allows for a number of improvements, including using more of the collected data (so that less information is sacrificed), controlling the effects of outliers, and allowing for the calculations of meaningful standardized effect sizes.

The mechanisms by which this is accomplished are relatively straightforward.

In the PND calculation, a proportion is generated from the percentage of data points exceeding the criterion: the highest baseline value. The PEM, using the *median in the baseline* as the criteria, allows the calculation of *four* proportions:

1. Calculate the proportion *above and below* the median in the *baseline phase* (Figure 4)
   Example results:
   Above = 50%
   Below = 50%
2. Calculate the proportion *above and below* the baseline median in the *intervention phase:*
   Example results:
   Above = 100%
   Below = 0%
3. Use the resulting ratios as the four-square elements in calculating a chi-square statistic, as described next.

When comparing the two ratios, the baseline ratio (at 50/50) is the expected value if the data were randomly distributed. If the intervention has no effect, the result would look very much like the baseline data. Any differences between the two, then, are a *goodness of fit* index, or how well the observed data (Intervention Phase) matches the expected data (Baseline Phase). This is the basis of a chi-square test: one of the most common and versatile nonparametric statistics. In fact, even

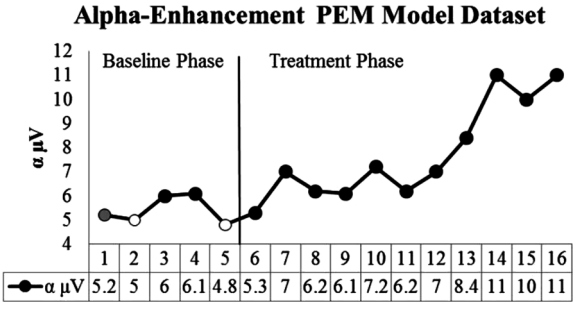


FIGURE 4. Percentage Exceeding Median (PEM) calculation method utilizing the model data set. *Note.* This approach measures the *proportion* of data points above and below the baseline median (gray marker) in both the baseline and treatment phases. Data points above the median are black; data points below the medium are white.

TABLE 1. Percentage of Data Points Exceeding the Median Crosstabs for Hypothetical Baseline Phase (A) and Intervention Phase (B) Data

	Phase A	Phase B	Totals
Proportion above median	50	80	130
Proportion below median	50	20	70
Totals	100	100	

in small data sets, Fisher's Exact Test can be used (Hintze, 2004), or Barnard's Test for the empirically intrepid (Barnard, 1945). These results can yield *Z* scores and significance values, which is an advantage not held by PND.

The Split-Middle Method and PEM can also produce effect size statistics from the two (baseline phase proportions) by two (intervention phase proportions) table calculated from the data. For example, Table 1 shows crosstabs for an experiment run where 80% of the intervention phase data were above the baseline median.

From these crosstabs, the calculation of a chi-square is a standard process. The chi-square, in turn, allows for the calculation of phi, a standard effect size statistic (Cohen, 1988) that would allow the results of the study to be included in meta-analyses, and a probability score based on Fisher's Exact Probability Test (Figure 5). There are a number of factors that influence the interpretation of the phi coefficient, including effect size, but a general interpretive guideline is: small effect size = .10, medium effect size =.30, and a large effect size = .50. There are many resources for these standard calculations, including a web-based calculator from Vassar College: see http://faculty.vassar.edu/lowry/tab2x2.html.

IRD

The IRD (Parker & Hagan-Burke, 2007) is a metric that appears to be particularly germane to psychophysiological research. It is closely associated with the procedural approaches used in PAND and PEM, is a widely accepted effect size metric, and is sensitive to changes in an individual's skill in producing changes

	Baseline (Phase A)	Intervention (Phase B)	Totals
Proportion Above Median	50	80	130
Proportion Below Median	50	20	70
Totals	100	100	200

	Chi-Square
Phi	Pearson
+.031	19.78
P	<.0001

	Fisher Exact Probability Test:	
P	one-tailed	0.0000069445853451661625
	two-tailed	0.000013889170690333838

FIGURE 5. Calculation of effect size statistics (e.g., chi-square, phi, Fisher's Exact Probability Test) using Percentage Exceeding Median (PEM) Crosstabs from Table 1. *Note.* The data analyzed are from the model data set.

in functional behavior. It has been used in many medical research studies under the name of *Risk Reduction* or *Risk Difference*.

The primary conceptual difference between IRD and other methods described in this review is that the IRD is calculated as the improvement rate in the intervention phase minus the improvement rate in the baseline phase, as opposed to the difference between phase averages. The primary distinction is that IRD accounts not only for average differences in treatment conditions but also trends within those conditions. Note that in the discussion of this calculation, the terminology ''treatment phase'' is used in place of ''intervention phase'' to remain consistent with the standard nomenclature of the formulae.

In many cases, the results of IRD calculations are correlated with PND, PAND, and PEM but have better discriminability than PND and can show negative effects for treatments that have iatrogenic outcomes (Parker et al., 2009). Moreover, in a series of studies, IRD was shown to be more strongly validated than PEM or PND and to have greater resilience with high variability data sets (Parker & Hagan-Burke, 2007).

Simply put, IRD is the difference between two improvement rates, which are expressed as two proportions. The proportions are generated in a fashion similar to PEM and PAND. The formula for calculating IRD is

$$IRD = IR_T - IR_B$$

Where:

$$IR_T = \frac{N\ improved\ data\ treatment\ phase}{Total\ data\ points\ treatment\ phase}$$

$$IR_B = \frac{N\ improved\ data\ points\ baseline\ phase}{Total\ data\ points\ points\ baseline\ phase}$$

The number of improved data points is defined as the points of data that *either tie or exceed* all of the data points in a contrasting phase (Figure 6 and Table 2). If you are up-training a variable, the IR_B would be all the data points in the baseline that are the same as or greater than any point in the treatment phase. Conversely, improved data points in the treatment phase are calculated as the number of data points that are greater than all of the data points in the baseline phase. The calculation method using a model data set is as follows:

1. Identify the number of nonoverlapping and total data points for each phase, as shown in the formula just presented.
2. The criteria for Improved Data Points are different for Baseline and Treatment Phases:
   a. Baseline = all data points *at or above any* data points in the Treatment Phase.

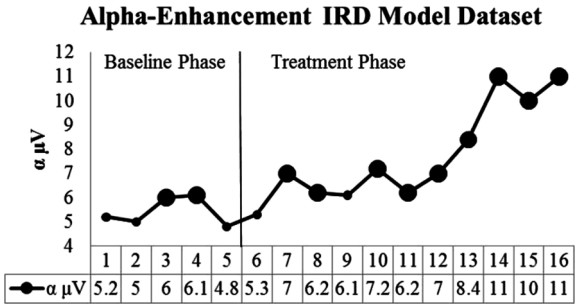


FIGURE 6. Improvement Rate Difference (IRD) calculation method utilizing the model data set. *Note.* IRD measures the amount of improvement and changes in direction from baseline in clinical trials. Improved Data Points in each condition are noted by large markers.

TABLE 2. Total Improved and Nonimproved Data Points in Baseline and Treatment Phases

Condition	Baseline	Treatment
Improved	2	9
Nonimproved	3	2
Total	5	11

b. Treatment Phase = all data points *above all* data points in the Baseline phase.

3. Calculate the Improvement Rate Ratio for each phase. Note again the differences between Baseline and Treatment Phases:
   a. Baseline: Improved data points are the overlapping data points:

$$\frac{Improved}{total} = \frac{2}{5} = 0.4$$

   b. Treatment: Improved data points are the nonoverlapping data points:

$$\frac{Improved}{total} = \frac{9}{11} = 0.82$$

4. Calculate

$$IRD = IR_T - IR_B = .82 - .4 = .42$$

5. Multiply the proportion by 100 to obtain a percentage: $.42 \times 100 = 42\%$
6. The interpretive statement for this result reads as ''The gain in improvement rate from the Baseline Phase to the Treatment Phase is a 42% gain, which is not a significant level of change.''

As with PEM and PAND, the interpretations of effect size are based on ratios obtained by comparisons of Baseline and Intervention Phase changes. An IRD of 50% (.50) indicates a chance level of difference, resulting in no improvement. The maximum value for improvement is equal to 1.0, and for iatrogenic effects the maximum decrease is −1.0 (Parker et al., 2009). Visual inspection and other empirical approaches interpret the model data set as significantly improved, but the IRD classifies it as not significant. It can be seen that an extension of the treatment phase would have been necessary to generate sufficient data to show clinically significant improvement. This is an added level of stringency that can be useful if the experimenter is interested in making causative statements.

Further refinement of interpretation can be generated with confidence intervals. Every basic statistical software package will have a function to generate confidence intervals. The confidence interval brackets can provide an index of how much faith, or confidence, to put into a set of results. Wide brackets indicate low precision and unreliable data, whereas tight brackets indicate the opposite. The 95% confidence interval is traditional for reporting in both research and applied settings.

To aid in the calculations and interpretations of IRD from study data, there are several no-cost resources digitally available. WinPEPI (Llorca, 2002) is available for download from http://www.brixtonhealth.com/pepi4windows.html, and Buchan (2006) provides a simple on-line program (http://www.phsim.man.ac.uk/).

HLM

A separate and very useful analytical tool for both analyzing data within a study and making that study more readily available for subsequent meta-analyses is the HLM (Raudenbush & Bryk, 2002). This analytical approach employs a multilevel approach to single-case designs, where Level 1 measures the change from baseline to intervention phase. Level 2 analyses accumulate data over multiple Level 1 studies and can be used to examine causal elements behind treatment effectiveness by isolating and accounting for different factors (Jenson et al., 2007). In other words, Level 1 analyses will typically be used with individual single-case design studies, whereas Level 2 analyses are employed in meta-analytic studies.

Although the calculation method is more complex, there are a number of advantages to this approach. The most apparent is that with a nested, hierarchical approach to analysis, it is a natural metric for inclusion in

sophisticated meta-analyses that may assist in isolating treatment component effectiveness over time. This calculation also demonstrates good control for Type I and Type II error rates and can generate *p* values and effect sizes with samples as small as 15 observations (Jenson et al., 2007).

The calculation for HLM Level 1 is

$$\gamma_T = \beta_0 + \beta_1(X_1) + e_T$$

Where:

γ_T = A participant's score at time T. If T = 1, it is the first observation in the first Baseline Phase.
β_0 = Mean of the first baseline phase
β_1 = Difference between the baseline mean and intervention mean
X_1 = A marker for which phase an observation is in. Traditionally, 0 = Baseline and 1 = Intervention.
e_T = error component

When this formula is examined, some features emerge: all of the data for one subject are summarized into two scores, which allow for direct comparisons across studies of differing observation lengths. In addition, because β_1 is a difference score, the summed scores allow for effect size calculations, similar to the IRD.

The calculation for HLM Level 2 is

$$\beta_{0j} = \gamma_{0o} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{1o} + \mu_{1j}$$

Where:

β_{0j} = The calculated baseline mean for each subject:
0 = Baseline
j = Participant number
γ_{0o} = Grand mean across subjects
μ_{0j} = Difference between the raw baseline mean for one subject (j) and the grand mean

And:

β_{1j} = The calculated difference between baseline and intervention phases, comprised of:
γ_{1o} = A subject's difference between Baseline and Intervention phase means
μ_{1j} = The difference between individual difference scores and the grand mean intervention difference score.

In both Level 1 and Level 2 analyses, calculation of effect sizes and significance levels are performed by a simple chi-square analysis, similar to other analyses covered in this article. Unlike the other analyses presented in this article, HLM is too complex to be efficiently calculated by hand, especially for larger or more complex data sets. There are programs designed specifically for this analysis that calculate all the requisite components (Raudenbush, Bryk, Cheong, & Congdon, 2004). This program does entail some cost but may be licensed for six months at a time for a reasonable fee. In addition, more universal statistical software packages are capable of doing HLM analyses including Statistical Package for the Social Sciences (SPSS: An IBM Company), and Statistical Analysis Software (SAS Institute Inc.), both of which are commercially available at varying licensing levels.

CONCLUDING REMARKS

Part 1 of this two-part series on single-case study design in psychophysiological research discussed the context, structure, and general techniques available to practicing clinicians and clinical scientists to generate evidence-based psychophysiological data. Utilization of these techniques makes it possible for the practice-level clinician to make meaningful contributions to the clinical literature, especially in the formulation of validated, empirically supported treatments. The techniques and design structures of single-case design are a substantive alternative to large sample size, randomized control trials, which are sometimes not the best research paradigm for clinical psychophysiology.

In this article, statistical techniques designed specifically for nonnormally distributed data were discussed within a context of single-case study designs in order to decrease barriers to utilizing these research methods. There is a range of simple analytic techniques that address issues of serial dependency that are not covered in this article, but are readily available in the scientific literature, that may be used with a broad array of research questions.

There are no ''hard and fast'' rules regarding which analyses to use for any given experimental procedure because the modular nature of the methodology. A general guideline, however, is that the more complex the design, the more sophisticated the analysis. For example, a simple A/B design can be sufficiently addressed with a PND analysis, whereas the inclusion of a withdrawal phase (such as recording without feedback) would benefit by PAND, being slightly more sophisticated. A multiple baseline with reversal phase would fit best with IRD or HLM analyses, and so on. Whenever possible, include analyses that provide an effect size metric, which will qualify the study for potential inclusion in later meta-analytic studies.

Ultimately, utilization of single-case design in psychophysiological research will maximize the accessibility of clinical data in rigorously designed empirical studies, which will allow clinical researchers to identify the efficacy and effectiveness of clinical interventions. Moreover, the inclusion of individual studies in larger meta-analytic reviews will help determine the most powerful elements of a clinical intervention and to cumulatively extend the depth and breadth of knowledge within the clinical literature.

REFERENCES

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Barnard, G. A. (1945). A new test for 2 × 2 tables. *Nature*, *156*, 177. doi:10.1038/156177a0

Borckardt, J. J., & Nash, M. R. (2002). How practitioners (and others) can make scientifically viable contributions to clinical-outcome research using the single-case time-series design. *International Journal of Clinical and Experimental Hypnosis*, *50*, 114–148. doi:10.1080/00207140208410095

Buchan, I. (2006). StatsDirect (Version 2.5.5) [Computer software]. Cheshire, UK: StatsDirect.

Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, *28*, 234–246. doi:10.1177/0145445503259264

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Franklin, R. D., Gorman, B. S., Beasley, T. M., & Allison, D. B. (1996). Graphical display and visual analysis. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 119–158). Mahwah, NJ: Erlbaum.

Gustafson, S. A., Nassar, S. L., & Waddell, D. (2011). Single-case design in psychophysiological research: Part I: Context, structure and techniques. *Journal of Neurotherapy*, *15*, 1–17.

Hays, W. L. (1994). *Statistics*. Fort Worth, TX: HarcourtBrace College.

Hintze, J. (2004). *NCSS and PASS [Computer software]*. Kaysville, UT: Number Cruncher Statistical Systems.

Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence-based practice, meta-analyses, and single subject designs. *Psychology in the Schools*, *44*, 483–493. doi:10.1002/pits.20240

Jones, W. P. (2003). Single-case time series with Bayesian analysis: A practitioner's guide. *Measurement and Evaluation in Counseling and Development*, *36*, 28–39.

Kavale, K. A., Mathur, S. R., Forness, S. R., Quinn, M. M., & Rutherford, R. B. (2000). Right reason in the integration of group and single-subject research in behavioral disorders. *Behavioral Disorders*, *25*, 142–157.

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

Kazdin, A. (1984). Statistical analysis for single-case experimental designs. In D. H. Barlow & M. Hersen (Eds.), *Single case experimental designs* (2nd ed., pp. 285–324). New York, NY: Pergamon.

Llorca, J. (2002). Computer programs for epidemiologists.PEPI v. 4.0. *Journal of Epidemiology Community Health*, *56*, 959–960. doi:10.1136/jech.56.12.959-c

Ma, H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification*, *30*, 598–617. doi:10.1177/0145445504272974

Manolov, R., & Solanas, A. (2008). Comparing $N=1$ effect size indices in presence of autocorrelation. *Behavior Modification*, *32*, 860–875. doi:10.1177/0145445508318866

Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy*, *34*, 189–211. doi:10.1016/S0005-7894(03)80013-8

Parker, R. I., & Hagan-Burke, S. (2007). Median-based overlap analysis for single case data: A second study. *Behavior Modification*, *31*, 919–936. doi:10.1177/0145445507303452

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*, *40*, 194–204. doi:10.1177/00224669070400040101

Parker, R. I., Vannest, K., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, *75*, 135–150. Retrieved from http://www.cec.sped.org/Content/NavigationMenu/Publications2/ExceptionalChildren/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Scruggs, T. E., & Mastropieri, M. A. (1994). The utility of the PND statistic: A reply to Allison and Gorman. *Behaviour Research and Therapy*, *32*, 879–883. doi:10.1016/0005-7967(94)90169-4

Scruggs, T. E., & Mastropieri, M. A. (2001). How to summarize single participant research: Ideas and applications. *Exceptionality*, *9*, 227–244.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *RASE: Remedial & Special Education*, *8*, 24–33. doi:10.1177/074193258700800206

Strain, P. S., Kohler, F. W., & Gresham, F. M. (1998). Problems in logic and interpretation with quantitative syntheses of single-case research: Mathur and colleagues (1998) as a case in point. *Behavioral Disorders*, *24*, 74–85.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2nd ed.). Columbus, OH: Merrill.