Abstract Title Page
Not included in page count.

Title: Effect size measure and analysis of single subject designs

Authors and Affiliations:

Hariharan Swaminathan, University of Connecticut (hariharan.swaminathan@uconn.edu)
H. Jane Rogers, University of Connecticut (jane.rogers@uconn.edu)

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Single-case research (SCR) methods provide a scientifically rigorous approach for documenting experimental control and advancing effective practices in education, behavior analysis and psychology (Kennedy, 2005; Kratochwill & Levin, 1992; McReynolds & Kearns, 1983; Richards, Taylor, Ramasamy & Richards, 1999; Tawney & Gast, 1984; Todman & Dugard, 2001.) Despite its long history and its potential for providing strong evidence for causal inference (Shadish & Rindskopf, 2007), single-case research has not collectively advanced evidence-based practice in these areas. Part of the reason can be traced to the fact that no acceptable statistical procedures are available for the analysis of single subject designs (SSD) and no satisfactory mechanism exists for synthesizing the results from single-case research studies.

The uses of single-case research are now expanding, however, and it is necessary to supplement visual analysis with a more formal index of effect size (Parker & Brossart, 2003; Parker et al., 2005). The movement to conduct meta-analyses with a body of research to document the basic patterns and findings in a field of study has begun to take hold in single subject research. The current commitment to documenting "evidence-based practices" (Kratochwill & Stoiber, 2002; Odom et al., 2005; Whitehurst, 2003) further emphasizes the need to move beyond statements based on results from any one study, toward documentation of effects supported by an integrated set of research results. Without a meaningful effect size index (Hedges, 2007) such syntheses of research and documentation of effective practices cannot take place (Horner, Swaminathan, Sugai, & Smolkowski, 2012).

Arguably, the most fruitful application of statistical procedures to single-case results may lie with developing strategies for measuring effect size. Effect size measures the magnitude of the difference in results in standardized units. More than 40 approaches for assessing effect size with single-case research have been proposed (Parker & Hagan-Burke, 2007; Kirk, 1996), and more recent proposals have become increasingly elegant (Parker, Hagan-Burke & Vannest, 2007). At this time, however, each of the available approaches carries important shortcomings (Maggin, Swaminathan, Rogers, O'Keefe, & Sugai, 2011). Measuring the size of experimental effects in single-case research will become a necessary element for the broad application of single-case results. To this end, an appropriate effect size measure has to be developed (Beretvas & Cheung, 2008; Hedges, 2007; Shadish & Rindskopf, 2007; Shadish & Rindskopf, & Hedges, 2008).

In addition to the problem of defining an index of effect size, the major concern raised by single-case researchers is the applicability of statistical procedures that fail to take into account the realities of single-subject designs. Serial dependence refers to the extent to which the patterns in repeated observations are functions of time separation between a series of values. While initially researchers (Huitema, 1985) expressed skepticism regarding the existence of serial dependence, researchers including Huitema and colleagues (Huitema, 1986; Huitema &McKean, 2007; McKnight, Mckean, & Huitema, 2000) have come to realize that serial dependence is an

importance issue that must be taken into account in the statistical analysis of single-subject design and have provided procedures that take into account serial dependencies in the data. Shadish et al. (2008) have emphasized this fact in their discussion of the methodological issues surrounding the analysis of single-subject designs.

Purpose / Objective / Research Question / Focus of Study:
Description of the focus of the research.
One of the vexing problems in the analysis of SSD is in the assessment of the effect of intervention. Serial dependence notwithstanding, the linear model approach that has been advanced involves, in general, the fitting of regression lines (or curves) to the set of observations within each phase of the design and comparing the parameters of these lines (or curves). In the simplest case of an AB design, this involves fitting a regression lines to the observations in phases A and B and comparing the slopes and intercepts. In the event the slope is zero in the two phases, the comparison reduces to the comparison of the means or levels. If there is a trend but the slope is the same in the two phases, then the adjusted means (levels) are compared. A problem arises if the levels and the trends in the two phases are different. In this case the intercepts and slopes are compared across the phases. While statistically such comparisons are not problematic, the interpretations of treatment effects that arise from these comparisons are not intuitive. A "combination" of the slope and intercept parameters that yields a single measure of treatment effect could be more meaningful. Such a measure of treatment effect yields, in turn, an effect size that can be readily computed and interpreted.

To this end, the study deals with:

1. Development of a measure of the effect of intervention which leads to an effect size measure
2. Development of statistical procedures for assessing intervention effects that take into account the serial dependence in single-subject designs;

The proposed statistical procedures are based on linear models, take into account the serial dependency in the observations, and are applicable when there is a mean shift as well as trend. In addition, an effect size measure that lends itself to meta-analysis is proposed.

Setting:
Description of the research location. s
(May not be applicable for Methods submissions)
Not Applicable
Population / Participants / Subjects:
Description of the participants in the study: who, how many, key features, or characteristics.
(May not be applicable for Methods submissions)
Not Applicable
Intervention / Program / Practice:
Description of the intervention, program, or practice, including details of administration and duration.
(May not be applicable for Methods submissions)
Not Applicable

Significance / Novelty of study:
Description of what is missing in previous work and the contribution the study makes.

As described in the previous section, this study provides a measure of treatment effect in single-subject designs, and consequently an effect size measure which could be used effectively in meta-analysis of single subject designs (Maggin, Swaminathan, Rogers, O'Keefe, & Sugai, 1991). In addition, analytical procedures, classical as well as Bayesian, that take into account serial dependency in the observations for testing hypotheses regarding the effect of intervention is provided. Furthermore, when multiple subjects are available, a D type effect size estimator is obtained employing the within and between subject variation.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

In the simplest case of an AB design, the linear models fitted to Phases A and B are:

$$y_t^A = \beta_0^A + \beta_1^A t + e_t \quad (t = 1, 2...., n_A)$$

$$y_t^B = \beta_0^B + \beta_1^B t + e_t \quad (t = n_A + 1, n_a + 2, ..., n_A + n_B)$$

If we set $\beta_1^A = \beta_1^B = 0$, we obtain a model where there is only a level change between the phases; on setting $\beta_1^A = \beta_1^B = \beta_1$ we obtain the model where there is a shift in the levels adjusted for the trend. Following the logic provided in Shadish, Cook, & Campbell (2002) in regards to quasi experimental time series designs, we conclude that if the trend lines in the two phases coincide, the intervention is not effective. This the effect of the intervention may be defined as the difference between the pints on the trend line in Phase B and the points of the Phase A trend line that is *projected* into Phase B. At point any point $n_A + t$ in phase B, the difference between the trend line in Phase B and the projected trend line from Phase A is

$$\delta_{AB(t)} = \hat{y}_{B|B(t)} - \hat{y}_{B|A(t)} = (\beta_0^B - \beta_0^A) + (\beta_1^B - \beta_1^A) * (n_A + t)$$

This difference is calculated at all points in Phase B and averaged to yield the treatment effect defined as

$$\Delta_{AB} = (\beta_0^B - \beta_0^A) + (\beta_1^B - \beta_1^A) * \frac{(2n_A + n_B + 1)}{2} \quad .$$

The estimate of the treatment effect measure defined above is obtained by substituting the estimates of the regression coefficients. It follows that if $\Delta = 0$ we can conclude that the intervention is not effective. To test the hypothesis $\Delta = 0$ (against the alternative, $\Delta \neq 0$), the ratio $D / SE(D)$ is formed and has an approximate *t*-distribution with degrees of freedom (N - 5), in the presence of first order autoregressive errors.

In the estimation of the regression coefficients and consequently D, and in particular, the standard error of *D, the* serial dependency in the observations must be taken into account. Maximum Likelihood and the Cochran-Orcutt procedures are classical approaches that can be employed to estimate the autocorrelation parameter, the regression coefficients, the error variance, and finally, *D*. Asymptotic as well as conditional standard errors are computed for testing the hypothesis $\Delta = 0$ and for constructing confidence intervals around Δ.

The parameter Δ is clearly an effect size measure. A standardized effect size measure is obtained by dividing by the standard deviation obtained by pooling the error variances in the two phases. An approximate distribution of the standardized effect size measure is obtained in this study.

An alternative and a more fruitful approach is to employ a Bayesian procedure to obtain the posterior distribution of the regression coefficients and Δ_{AB}, the treatment effect parameter. The advantage of this approach is that the posterior distribution contains all the information needed for drawing inferences. By computing the $\frac{1}{2}\alpha$ and $(1-\frac{1}{2}\alpha)$ percentile points, $(1-\alpha)$% credibility intervals can be constructed without making any distributional assumptions. Markov-Chain Monte Carlo (MCMC) procedure as implemented in WINBUGS is employed for obtaining the Bayesian estimates.

In the event several subjects are included in the study as in multiple baseline designs, a Bayesian procedure that in principle equivalent to hierarchical linear modeling approach, is proposed in the study. This approach takes into account the intra and inter subject variation to provide a measure of variability that yields a D-type effect size estimator. This approach is illustrated using the MCMC procedure as implemented in WINBUGS.

Usefulness / Applicability of Method:
Demonstration of the usefulness of the proposed methods using hypothetical or real data.

The methods described above are applied for the analysis of several standard sets of data so that the different approaches proposed by the symposium presenters for analyzing single subject designs can be compared. Classical and Bayesian procedures are compared.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

The statistical procedure developed here that includes a definition of treatment effect and consequently an effect size provides a viable method for the analysis of single-subject designs. Then advantage of the procedure outline here is that a single measure of treatment effect and consequently a single effect size measure is available when there is both a level change and trend change in the phases. Furthermore, the serial dependency in the observations is taken into account in the analysis of the data.

Single-subject designs are very short time series and this presents a problem in estimating the parameters in the autoregressive process assumed for the errors. The Bayesian approach described has the potential for overcoming this limitation.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Beretvas, S. N., & Chung, H., (2008,). *Computation of regression-based effect size measures.* Paper presented at the annual meeting of the International Campbell Collaboration Colloquium, Vancouver, Canada.

Beretvas, S. N., & Chung, ., (2008,). An evaluation of modified R^2 change effect size indices for single subject experimental designs *Evidence-Based Communication Assessment and Intervention*, 2:3,120-128.

Center, B.A., Skiba, R.J., and Casey, A. (1985-1986). A methodology for the quantitative synthesis of intra-subject design research. *Journal of Special Education, 19*, 387-400.

Cochrane, D. and Orcutt, G.H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association 44*, 32-61.

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

Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, *41*, 372-384.

Ferron, J. M., Farmer, J. L., Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel modeling approaches. *Behavior Research Methods*, *42*, 930-943.

Fisher, W., Kelley, M., and Lomas, J. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis 36,* 387-406.

Hedges, L. V. (2007). Meta-analysis. In C. R. Rao and S. Sinharay (Eds.) *The handbook of statistics.* Amsterdam: Elsevier.

Hedges, L. V., Shymansky, J. A., and Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis*. Washington, D.C.: National Science Teachers Association.

Horner, R., Swaminathan, H., Sugai, G., & Smolkowski, K. (in press). Expanding analysis and use of single-case research. *Education and Treatment of Children*.

Huitmea, B. (1985). Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment, 7*, 107-118.

Huitema, B. (1986). Autocorrelation in behavioral research: Wherefore art thou? In A. Poling and R.W. Fuqua (Eds.). *Research methods in applied behavior analysis: Issues and advances*. (pp. 209-223). New York: Plenum.

Huitema, B.E. (1990). Autocorrelation: 10 years of confusion. *Behavioral Assessment, 10*, 253-294.

Huitema, B., and McKean, J. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement, 60*(1). 38-58.

Huitema, B., and McKean, J. (2007). Identifying autocorrelation generated by various error processes in interrupted time-series regression designs: A comparison of AR1 and Portmanteau tests. *Educational and Psychological Measurement, 67*(3), 447-459.

Jones, R.R, Weinrott, M.R., and Vaught, R.S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. *Journal of Applied Behavior Analysis*,*11*, 278-283.

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn and Bacon.

Kratochwill, T. and Stoiber, K. (2002). Evidence-based interventions in school psychology: Conceptual foundations for the Procedural and coding manual of Division 16 and Society for the study of school Psychology Task Force. *School Psychology Quarterly, 17,* 341-389.

Lambert, M. C., Cartledge, G., Heward, W. L., & Lo, Y.-Y. (2006). Effects of response cards on disruptive behavior and academic responding during math lessons by fourth-grade urban students. Journal of Positive Behavior Interventions, 8, 88-99.

Maggin, D. Swaminathan, H. Rogers, H.J. ,O'Keefe, B., Sugai, G.,& Horner, R. (2011). A Generalized Least Squares Regression Approach for Computing Effect Sizes in Single-Case Research: Application Examples. *Journal of School Psychology*, 49, 301-321.

Matyas, T.A. and Greenwood, K.M. (1996). Serial dependency in single-case time series. In R.D. Franklin, D.B. Allison, & B.S. Gorman (Eds.). *Design and analysis of single-case research.* (pp. 215-243). Mahwah, NJ: Lawrence Erlbaum.

McKnight, S., Mckean, J.W., & Huitema, B.E. (2000). A double bootstrap method to analyze to analyze an intervention time series model with autoregressive error terms. *Psychological Methods, 5*, 171-178.

Moeyaert, M., Ugille, M., Ferron, J., Beretvas, T., & Van den Noortgate, W. (in press). Three-level analysis of single-case experimental data: Empirical validation. *The Journal of Experimental Education*.

Nugent, W.R. (1996). Integrating Single-Case and Group-Comparison Designs for Evaluation. The Journal of Applied Behavioral Science, 32, 209-226. DOI: 10.1177/0021886396322007

Odom, S.L., Brantlinger, E., Gersten, R., Horner, R.H., Thompson, B., andHarris, K. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children 71*(2), 137-148.

Parker, R., and Brossart, D. (2003). Evaluating single case research data: A comparison of seven statistical methods. *Behavior Therapy 34*, 189-211.

Parker, R., Brossart, D., Callicott, K., Long, J., Garcia De-Alba, R., Baugh, F., and Sullivan J. Effect size in single-case research: How large is large? (2005). *School Psychology Review 34*, (1), 116-132.

Parker, R., and Hagan-Burke, S., (2007). Useful effect size interpretations for single case research. *Behavior Therapy, 38*, 95-105.

Parker, R., Hagan-Burke, S., and Vannest, K. (2007). Percent of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education, 40* (4), 194-204.

Parsonson, B., and Baer, D. (1992). Visual analysis of data, and current research into the stimuli controlling it. In T. Kratchowill and J. Levin (Eds.) *Single-case Research Design and Analysis: New Directions for Psychology and Education.* (pp. 15-40). New Jersey: Lawrence Erlbaum Associates.

Perone, M. (1999). Statistical inference in behavior analysis: Experimental control is better. *The Behavior Analyst, 22*(2), 109-116.

Rogers, H.J. and Swaminathan, H. (2007). A computer program for the statistical analysis of single subject designs. Storrs, CT: University of Connecticut, Measurement, Evalauation, and Assessment Program.

Rosnow, R. L., and Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Pyschological Methods, 1,* 331-340.

Searle, R. (1971). *Linear Models.* New York: John Wiley.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Shadish, W., Rindskopf, D.M., & Hedges, L.V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence- Based Communication Assessment and Intervention,* 2:3, 188-196..

Shadish, W.R., Hedges, L.V., Pustejovsky, J., Rindskopf, D.M., Boyajian, J.G. & Sullivan, K.J. (in press). Analyzing single-case designs: *d*, *G*, multilevel models, Bayesian estimators, generalized additive models, and the hopes and fears of researchers about analyses. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-Case Intervention Research: Methodological and Data-Analysis Advances*. Washington, D.C.: American Psychological Association.

Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation*, *113*, 95-109.

Shadish, W.R. & Sullivan, K.J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods, 43*, 971-980. DOI 10.3758/s13428-011-0111-y. ERIC #ED530280.

Shavelson, R., and Towne, L. (2002) *Scientific Research in Education*. National Academy Press, Washington, D.C.

Swaminathan, H., and Algina, J. (1977). Analysis of quasi-experimental time-series designs. *Journal of Multivariate Behavioral Research*, *12*, 111-131.

Tawney, J.W. and Gast, D.L. (1984). *Single subject research in special education*. Columbus: Merrill.

Ugille, M., Moeyaert, M., Beretvas, T., Ferron, J., & Van den Noortgate, W. (in press). Multilevel Meta-Analysis of Single-Subject Experimental Designs: A Simulation Study. *Behavior Research Methods*.

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, *18*, 325-346.

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*, *35*, 1-10.

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*, 142-151.

White, D.M., Rusch, F.R., Kazdin, A.E. and Hartmann, D.P. (1989). Applications of meta-analysis in individual subject research. *Behavioral Assessment, 11*, 281-296.

Whitehurst, G.J. (2003). Evidence-based education. PowerPoint presentation. http://www.ed.gov/offices/OERI/presentations/evidencebase.ppt