Original Article

Research Synthesis Methods

Received 26 February 2012, Revised 11 June 2012, Accepted 25 June 2012 Published online 14 August 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1052

A standardized mean difference effect size for single case designs

Larry V. Hedges,[a] James E. Pustejovsky[a*†] and William R. Shadish[b]

Single case designs are a set of research methods for evaluating treatment effects by assigning different treatments to the same individual and measuring outcomes over time and are used across fields such as behavior analysis, clinical psychology, special education, and medicine. Emerging standards for single case designs have focused attention on the need for effect sizes for summarizing and meta-analyzing findings from the designs; although many effect size measures have been proposed, there is little consensus regarding their use. This article proposes a new effect size measure for single case research that is directly comparable with the standardized mean difference (Cohen's *d*) often used in between-subjects designs. Techniques are provided for estimating the new effect size, as well as its variance, from balanced or unbalanced treatment reversal designs. The proposed estimation methods are further evaluated using a simulation study and then demonstrated in two applications. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: single case designs; effect size; autocorrelation; hierarchical linear model

Single case designs are distinguished by the fact that they assign different treatments to the same individual and measure an outcome over time. Such designs are a special type of the broader category of repeated measures designs. These designs are widely used in behavior analysis, clinical psychology, and special education, and sometimes used in medicine. In principle, single case designs permit the detection of treatment effects from a study that involves a single case and a comparison between two periods (one that is a baseline or control period and another period in which a treatment is assigned to that case). In practice, studies using single case designs usually involve more than one period in each treatment condition and replications of the design with more than one case. Emerging standards for single case designs emphasize at least two periods of each treatment (leading to three treatment contrasts or reversals) and replications across several individuals. [15]

Evaluation of the results of single case designs involves the search for functional relations between treatment assignment and an outcome. That is, in the ideal, the study is designed so that each treatment (or baseline) phase is continued for enough measurements that the pattern of outcome values is clearly established. To establish functional relations, researchers prefer stability within treatment phases, with treatment effects conceived as differences in these stable patterns between treatment and control phases. Stability, however, can be conceptualized in many different ways. For example, the pattern could be one of fluctuation around a constant value with a common mean within a phase and a common residual variance within all phases. The pattern could also involve systematic increase or decrease across measurements in a phase, such as a linear or quadratic trend, and a common residual variance within phases. Alternatively, the pattern could include a constant mean or a trend over measurements accompanied by larger or smaller residual variation around the trend, depending on whether the treatment is in effect. From this perspective, functional relations between treatment and outcome (what one would call treatment effects in between-subjects designs) are understood to be contrasts between the stable states established within treatment phases.

Statistical analysis is not always used in assessing treatment effects in single cases designs. One reason for this is that repeated measurements from the same individual often cannot be considered independent (i.e. they involve an autocorrelation structure), which complicates statistical analysis. Another reason is that stability within phases may take many forms, each of which requires a somewhat different statistical analysis to evaluate

[a]*Northwestern University, Evanston, IL, USA*
[b]*University of California, Merced, CA, USA*
**Correspondence to: James E. Pustejovsky, Northwestern University.*
[†]*E-mail: pusto@u.northwestern.edu*

comparisons among phases. Still, once a particular conception of stability is posited, there is no principled reason that statistical methods (including hypothesis tests) cannot be used in the analysis of treatment effects in single case designs. For a review of statistical methods for such analyses, see Kratochwill and Levin. [16]

Treatment effects are quantified using measures called effect sizes, which are useful for a variety of reasons. Effect size measures are often used as a supplement to hypothesis tests and as a way to quantitatively summarize study results (see [25]). Effect size measures also have an important function in making more comparable the results of studies using different designs and outcome measures (see, e.g. [12,24]). In one approach to systematic research review, effect sizes provide the basis of formal quantitative syntheses (see, e.g. [7]). There is substantial consensus on methods for computing effect sizes in between-subjects designs, and many 'standard' effect size measures are well known among researchers (e.g. [5,8]). In contrast, there is much less consensus on methods for computing effect sizes in single case designs (see, e.g. [2,18,20,22,4]).

One reason that defining effect sizes in single case design is difficult is that they represent differences between stable patterns in different phases. Because there are many different kinds of stability, it is difficult to propose an effect size measure that is equally appropriate for expressing differences between phases for all of them. We believe that some progress can be made by focusing on specific types of stable data patterns and defining effect size measures that express the differences between phases for stable patterns of a single, given type.

The purpose of the present paper is to propose a new effect size measure and corresponding estimation techniques for single case designs. The focus is on a single type of stable pattern, arguably the simplest type: fluctuation around a constant value (a common mean with a common residual variance within phases). We offer a statistical model in which the effect size parameter corresponds to the standardized mean difference (Cohen's *d*), a well-known effect size parameter in between-subjects designs. Our effect size measure thus has the virtue of expressing the treatment effect from single case designs on the same metric as that often used in between-subjects designs. We propose an estimator for this effect size, derive its approximate sampling distribution (including expressions for the mean and variance), and evaluate the accuracy of the analytic expressions for the mean and variance of the estimator. The initial exposition describes the effect size in the context of a two-phase $(AB)^1$ design with equal numbers of time points within each phase, where A indicates one phase (e.g. baseline), B indicates a second phase (e.g. treatment), and the superscript indicates the number of AB sequences in the study. In a later section, we generalize the results to designs with $2k$ phases, so-called $(AB)^k$ designs, in which each individual may have an unequal number of time points in different phases.

1. The balanced $(AB)^1$ design with *n* observations in each phase

1.1. Model

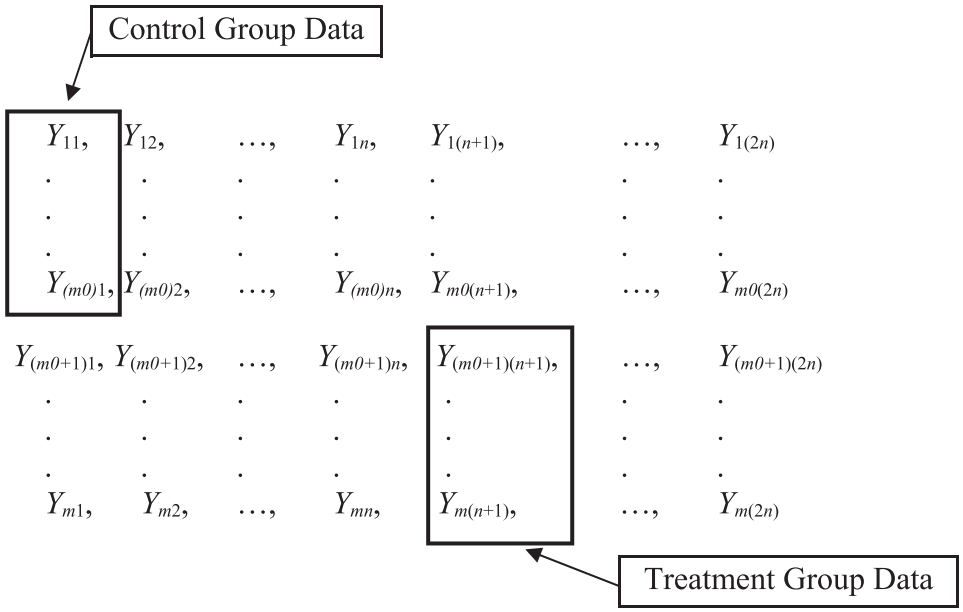
We begin by positing a structural model for the data collected from an AB design. This model is broad enough to encompass both a between-subjects experiment and a single case design with replications across individuals, making it possible to identify a parameter that is a conventional effect size (the standardized mean difference or *d*-index) in the between-subjects design. We then show that it is possible to estimate the same parameter using data collected with a single case design.

To simplify the initial exposition, consider a two-period $(AB)^1$ design. Let Y_{ij} be the j^{th} observation from the i^{th} individual, where there are $m > 2$ individuals, and for each individual, suppose that the first n observations are in the baseline period, followed by n observations in the treatment period. Thus, the data are denoted Y_{ij} for $i = 1, \ldots, m$ and $j = 1, \ldots, 2n$. We can describe the entire data layout as follows.

Control				Treatment			
Y_{11},	Y_{12},	…,	Y_{1n},	$Y_{1(n+1)}$,	$Y_{1(n+2)}$,	…,	$Y_{1(2n)}$
Y_{21},	Y_{22},	…,	Y_{2n},	$Y_{2(n+1)}$,	$Y_{2(n+2)}$,	…,	$Y_{2(2n)}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
Y_{m1},	Y_{m2},	…,	Y_{mn},	$Y_{m(n+1)}$,	$Y_{m(n+2)}$,	…,	$Y_{m(2n)}$

There is a between-groups experiment with total sample size m embedded within this data structure. Suppose that m_1 cases are randomly assigned to receive the treatment condition, the remaining m_0 cases receive the control condition, and a single observation is made from each case at time $j = 1$. We can adjust the numbering of cases so that the first m_0 cases ($i = 1, \ldots, m_0$) are the control condition and the second set of m_1 cases ($i = m_0 + 1, \ldots, m$) are in the treatment condition. This is shown by the two boxes in the illustration that follows. Note that because there is only one observation per case and it is reasonable to assume that cases are independent, the data should meet the usual assumptions invoked in the analysis of a between-subjects design. Further, note that

we have chosen one subset of the data that yielded a between-subjects experiment, but this is only one of many possibilities. For example, we might choose the first observation in each period, or the second, or the third, and so on. Similarly, we might choose to use the first observation in both periods, or we might choose the first in the control period but the second (or the third) in the treatment period, and so on.



We posit the following stochastic model for such data: (1) the Y_{ij} are normally distributed, (2) the data series for each case lacks any time trend, and (3) within each case, the deviations from the mean at each point in time are weakly stationary, with first-order autocorrelation ϕ. Specifically, the statistical model for the first period is

$$Y_{ij} = \mu^C + \eta_i + \varepsilon_{ij}, \qquad i = 1, \ldots, m; j = 1, \ldots, n$$

and the statistical model for the second period is

$$Y_{ij} = \mu^T + \eta_i + \varepsilon_{ij}, \quad i = 1, \ldots, m; j = n+1, \ldots, 2n$$

where $\mu^T - \mu^C = \Delta$ represents the shift between baseline and treatment periods. We assume that the individual level effects η_i are normally distributed with zero mean and variance τ^2. The assumption that each time series is weakly stationary implies that, conditional on the η_i, the covariance of Y_{ij} with $Y_{i(j+t)}$ depends only on t. We assume further that the ε_{ij} have variance σ^2 and first-order autocorrelation ϕ within individuals, so that

$$\mathrm{Cov}\{\varepsilon_{ij}, \varepsilon_{i'j'}\} = \begin{pmatrix} 0 & \text{if} \quad i \neq i' \\ \phi^{|j-j'|}\sigma^2 & \text{if} \quad i = i' \end{pmatrix}.$$

The $2n \times 2n$ covariance matrix of the errors within individuals therefore has the form

$$\begin{pmatrix} \sigma^2 & \phi\sigma^2 & \ldots & \phi^{2n-1}\sigma^2 \\ \phi\sigma^2 & \sigma^2 & \ldots & \phi^{2n-2}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{2n-1}\sigma^2 & \phi^{2n-2}\sigma^2 & \ldots & \sigma^2 \end{pmatrix}.$$

2. The effect size parameter

Because the variance of observations within cases is σ^2 and the variance of observations between cases is τ^2, the total variance of each observation is $\sigma^2 + \tau^2$. Under this model, define the effect size parameter

$$\delta = \frac{\mu^T - \mu^C}{\sqrt{\sigma^2 + \tau^2}} \tag{1}$$

This definition of the effect size is precisely the standardized mean difference (Cohen's *d*-index) that is widely used in between-subjects experiments. Note that the variance here consists of a within-person component σ^2 and a between-person component τ^2 that are not separable in a between-subjects design. Thus, the variance estimated in the between subjects design is $\sigma^2 + \tau^2$. Because the effect size parameter is the same in either the single case

design or a corresponding between-subjects design, estimates of this parameter will be on the same scale, and thus comparable in magnitude, whether they are based on data from a single case design or a between-subjects design.

3. Estimation of effect size

There are several possible approaches to estimation of δ. The approach used in this paper is to compute a pooled (across cases) estimate of $(\mu^T - \mu^C)$ and a pooled (across cases) estimate of $(\sigma^2 + \tau^2)$ and to combine those estimates to obtain an estimate of δ.

Other approaches to the estimation of δ are certainly possible. One approach is to estimate $(\mu^T - \mu^C)$, σ, and τ individually (e.g. using maximum likelihood) and combine the estimates to obtain an estimate of δ. This approach has the disadvantage that the small sample properties of the combined estimate of δ are difficult to obtain analytically; consequently, there is no obvious method to determine the bias of the resulting estimate, so that it can be reduced or eliminated. A second approach is to note that comparing single points in time in the control and treatment phases (the implicit between-subjects study) can yield an estimator that is a conventional standardized mean difference, one whose sampling distribution is known. A disadvantage of that estimator is that it throws away most of the data. A third approach is to average a series of estimates using the second approach (e.g. one for each time point). This has the advantages that it uses all the data and has known (but rather complicated) multivariate *t*-distribution theory, but it has the disadvantages that the average has a very complex distribution. Also, we found in preliminary, unreported simulation work that this estimator is much less efficient than others; that is, its sampling distribution has larger variance than other approaches.

The approach presented in this paper is to compute pooled estimates of $(\mu^T - \mu^C)$ and $(\sigma^2 + \tau^2)$ and to use these estimates to obtain an estimate of δ. This is similar to approaches two and three. It has the advantage of providing direct estimates of the numerator and denominator of δ. Moreover, the numerator has a normal distribution, and the denominator is approximately the square root of a random variable having a chi-squared distribution; the ratio therefore has a distribution that is approximately proportional to a non-central *t*-distribution, just as in the case of between-subjects experiments. This approach also has a less obvious advantage. The bias and to some extent variance of the effect size estimate depend on the effective number of degrees of freedom in the denominator (more is better). Whereas approaches two and three have a denominator with $m - 1$ degrees of freedom (one less than the number of individuals), the approach used here has a denominator whose sampling distribution has more than $m - 1$ degrees of freedom and thus has improved sampling properties. Some rather extensive simulation studies confirmed that the approach used in this paper has more desirable sampling properties (less bias, more accurate variance approximations, and smaller mean squared error) than the three alternatives mentioned earlier.

Define the effect size estimate ES via

$$\mathrm{ES} = \frac{\bar{D}}{S} \tag{2}$$

where

$$\bar{D} = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{n}\sum_{t=n+1}^{2n} Y_{it} - \frac{1}{n}\sum_{t=1}^{n} Y_{it}\right) \tag{3}$$

$$S^2 = \frac{1}{2n(m-1)}\sum_{t=1}^{2n}\sum_{i=1}^{m}(Y_{it} - \bar{Y}_{\bullet t})^2 \tag{4}$$

and $\bar{Y}_{\bullet t}$ is the mean across individuals at the t^{th} time point given by

$$\bar{Y}_{\bullet t} = \frac{1}{m}\sum_{i=1}^{m} Y_{it}.$$

It follows that $\bar{D}$ is an unbiased estimator of $(\mu^T - \mu^C)$, and S^2 is an unbiased estimate of the total variance $(\sigma^2 + \tau^2)$.

Under this model, the variance of $\bar{D}$ is

$$\mathrm{V}\{\bar{D}\} = \frac{2(b_1 - c_1)\sigma^2}{m} \tag{5}$$

where b_p and c_p are functions of the autocorrelation ϕ and the phase length n, defined in general as

$$b_p = \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\phi^{p|s-t|} = \frac{1}{n} + \frac{2}{n^2}\sum_{t=1}^{n-1}\phi^{pt}(n-t) \tag{6}$$

and

$$c_p = \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\phi^{p|n+t-s|} = \frac{1}{n^2}\sum_{t=1-n}^{n-1}\phi^{p(n+t)}(n-|t|). \tag{7}$$

Here, we define b_p and c_p in general because, although the variance of $\bar{D}$ depends only on b_1 and c_1 (that is b_p and c_p for $p=1$), we will need b_2 and c_2 in expressions for the variance of S^2. Under the aforementioned model, the variance of S^2 is

$$\mathrm{V}\{S^2\} = \frac{\left[(b_2+c_2)(1-\rho)^2 + 2(b_1+c_1)\rho(1-\rho) + 2\rho^2\right](\sigma^2+\tau^2)^2}{m-1} \tag{8}$$

where

$$\rho = \frac{\tau^2}{\tau^2+\sigma^2} \tag{9}$$

is a kind of intraclass correlation that represents the between-person variance τ^2 as a fraction of the total variance $(\tau^2+\sigma^2)$.

Using a theorem in Box [6] on the distribution of quadratic forms in normal variables, it follows that the sampling distribution of S^2 is approximately a chi-squared with ν degrees of freedom, where ν is given by

$$\nu = \frac{2(m-1)}{(b_2+c_2)(1-\rho)^2 + 2(b_1+c_1)\rho(1-\rho) + 2\rho^2}. \tag{10}$$

Therefore, ES is a constant θ times a random variable with the non-central t-distribution with ν degrees of freedom, where θ is given by

$$\theta = \sqrt{\frac{\mathrm{V}\{\bar{D}\}}{\tau^2+\sigma^2}} = \sqrt{\frac{2(b_1-c_1)(1-\rho)}{m}} \tag{11}$$

It follows from the results in Hedges [9] that the bias in ES can be corrected by multiplying ES by the factor

$$\mathrm{J}(\nu) = 1 - \frac{3}{4\nu-1}. \tag{12}$$

so that the effect size

$$G = \mathrm{J}(\nu)\mathrm{ES} \tag{13}$$

is approximately an unbiased estimator of δ.

It also follows that the variance of G is approximately

$$\mathrm{V}\{G\} = \mathrm{J}(\nu)^2\left[\frac{\nu\theta^2}{\nu-2} + \delta^2\left(\frac{\nu}{\nu-2} - \frac{1}{\mathrm{J}(\nu)^2}\right)\right] \tag{14}$$

A slightly simpler asymptotic approximation of the variance is

$$\mathrm{V_A}\{G\} = \mathrm{J}(\nu)^2\left[\theta^2 + \frac{\delta^2}{2\nu}\right]. \tag{15}$$

Note that the expressions (13) for G and (14) and (15) for its variance involve the nuisance parameters ϕ and ρ. In application, (14) or (15) will need to be evaluated using the estimate G in place of δ, along with estimates for the nuisance parameters.

4. Accuracy of the approximate sampling distribution

We investigated the sampling distribution of the estimator under two conditions: first, when the nuisance parameters ϕ and ρ were known and second, when the nuisance parameters were estimated. For each part, we limited consideration to the $(AB)^1$ design because this is the simplest design, although it has the essential

features of all $(AB)^k$ designs. Designs with more periods should have larger sample sizes; we would therefore expect the estimator to perform no worse (and probably better) in these designs. For each analysis, we varied ϕ and ρ over their entire parameter spaces. We considered values of n and m that appear to be representative of values found in the single case literature, based on a recent survey by Shadish and Sullivan. [23] Table 1 reports the parameter values and sample sizes used in the analyses.

First, we considered the bias of G and the accuracy of the expressions for its variance if known values of the nuisance parameters are used. For known ϕ and ρ, analytic expressions are available for the moments of G; we provide a derivation in Appendix A. Using these to evaluate the bias of G, we found it to be small, as expected. For example, when $m=4$ and $n=4$, the estimated bias never exceeds 0.04δ (relative bias of 4%). For more moderate values of $0<\rho\le 0.4$ and $-0.5\le\phi\le 0.5$, the bias is always less than 0.02δ.

The variance (14) is quite accurate when the intraclass correlation ρ is small to moderate (say below 0.3), but it tends to overestimate the variance somewhat as ρ becomes large, a tendency that is exaggerated when there is a large negative or positive autocorrelation. The simpler variance estimate (15) tends to underestimate the variance, often substantially. A typical result is shown in Fig. 1, which plots the estimated relative variance (the average value of the variance estimate divided by the exact value of the variance) on the vertical axis against ϕ on the horizontal axis. The effect size is fixed at $\delta=0.4$, and separate curves are plotted for $\rho=0.2$, 0.4, and 0.6; each of the panels in the figure corresponds to a different combination of values of m and n. In this figure, all of the curves above the relative variance of 1.0 correspond to the variance estimate (14), whereas those below a relative bias of 1.0 correspond to the simpler variance estimate (15). Note that the vertical scale of these graphs varies depending on the value of m.

To determine the properties of the effect size estimate G and its variance when the nuisance parameters are estimated, we carried out a simulation study. The simulation involved five parameters, the levels of which are reported in Table 1. The number of cases, m, was varied from 4 to 12 to capture a range of values observed in practice. The number of observations per period, n, assumed to be equal in the baseline and treatment periods, was varied from 4 to 12. The average treatment effect may range greatly depending on the treatment under study; we therefore examined a wide range of levels. We varied the autocorrelation ϕ and the intraclass correlation ρ over all but the most extreme possible values. The total variance $\tau^2+\sigma^2$ was fixed equal to one. For each combination of the parameters δ, ϕ, ρ, m, and n, we simulated 8000 iterations of the model. Rather than attempting to summarize the results across the entire set of simulation parameters, we report results for selected margins. Full simulation results, as well as the R code used to generate the simulations, are available from the second author.

For purposes of simulation, we used the simplest estimators of ϕ and ρ, the method-of-moments (sometimes called the Yule–Walker) estimators, with simple bias-reducing corrections; see Appendix C for details. These estimators are known to be biased, even after correction, when computed from single short time series (see, e.g. [14]). Other estimators are available from the literature on single time series, but the generalization of these estimators to unbalanced multiple time series is not straightforward, nor is it obvious whether, or how much, these generalizations might improve on simpler estimators of ϕ and ρ.

Our simulation studies when ϕ and ρ are estimated from the data confirmed that the bias of G remains small, except in the case of very large (and probably unrealistic) negative autocorrelations (e.g. $\phi=-0.9$). When $|\phi|\le 0.5$ and $\rho\le 0.5$, the relative bias of G is always less than 3% in absolute value. It appears that the variance of G is estimated more poorly when ϕ and ρ are estimated from the data, regardless of whether (14) or (15) is used. Figure 2 provides an illustration; it is constructed just as Fig. 1, plotting the relative variance (the average estimated variance divided by the true variance) as a function of ϕ for various values of m, n, and ρ. It is likely that improved estimation of ϕ and ρ may yield improved variance estimation; this remains a topic for future investigation.

5. The $(AB)^k$ design with unequal numbers of observations in each phase for each individual

Suppose that a study uses an $(AB)^k$ design, so that there are $2k$ phases, and m cases. We allow that each individual case may have a different number of observations in each phase. Let n_i^a, $a=1,\ldots,2k$, $i=1,\ldots,m$ be the number of observations in the a^{th} phase for the i^{th} individual, and define $n_i^0=0$ for all $i=1,\ldots,m$.

Table 1. Simulation parameter values

Parameter	Definition	Number of levels	Minimum	Step	Maximum
m	Number of cases	3	4	4	12
n	Observations per period	3	4	4	12
δ	Effect size	4	0.0	0.4	1.2
ϕ	Autocorrelation	10	−0.9	0.2	0.9
τ	$\tau^2/(\tau^2+\sigma^2)$	5	0.0	0.2	0.8

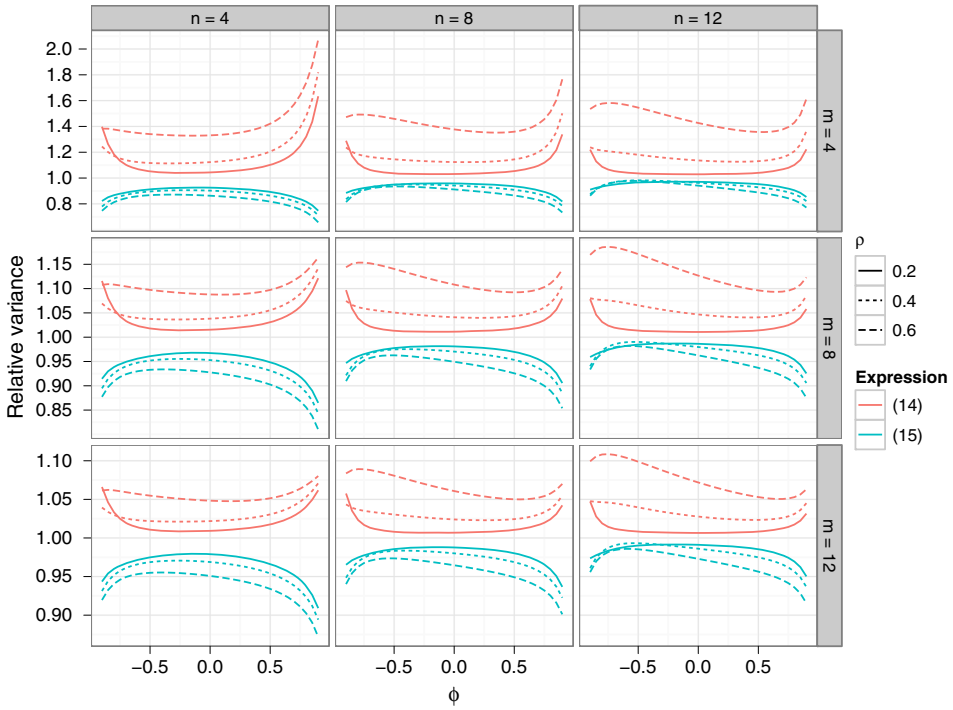


Figure 1. Ratio of variance computed from (14) and (15) to the exact variance of G when ϕ, ρ, and $\delta = 0.4$ are known.

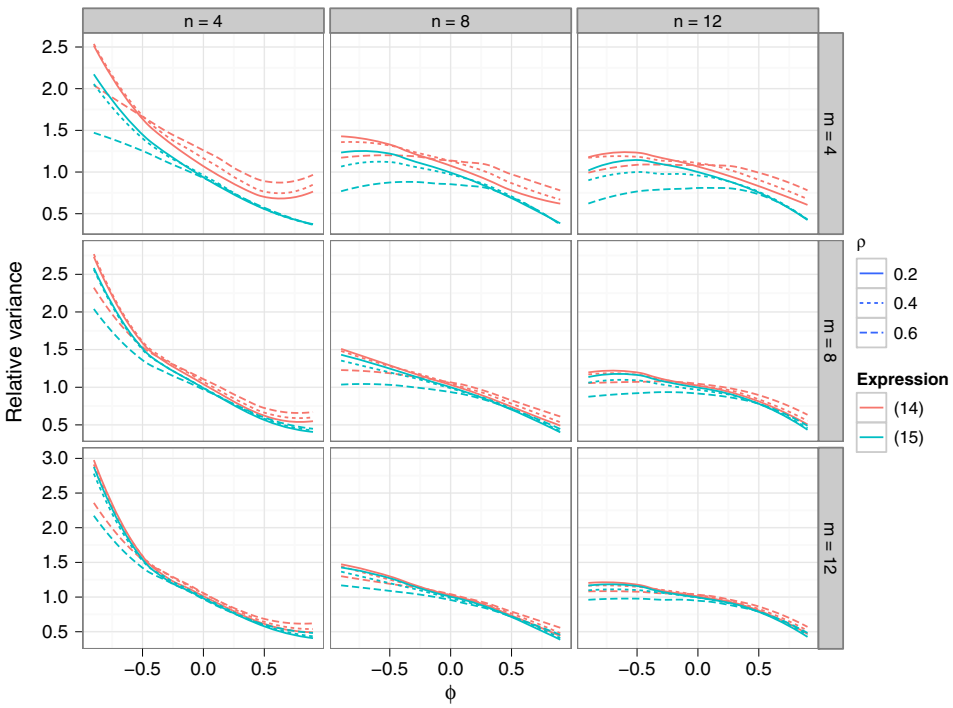


Figure 2. Ratio of the variance computed from (14) and (15) to the simulated variance of G when ϕ and ρ are estimated, for $\delta = 0.4$.

Because the number of observations within phases is not the same across individuals, we need a notation that can denote the first, second, and so on observation within each phase for each individual. Define the total number of observations (j-values) for the i^{th} individual through the a^{th} phase to be N_i^a, so that

$$N_i^a = n_i^0 + n_i^1 + \cdots + n_i^a$$

and define $N_\bullet^\bullet$ to be the total number of observations (the sum of the n_i^a) so that

$$N_{\bullet}^{\bullet} = \sum_{i=1}^{m} N_i^{2k} = \sum_{i=1}^{m} \sum_{a=1}^{2k} n_i^a.$$

The a^{th} phase for the i^{th} individual includes the j-values between $N_i^{a-1}+1$ and $N_i^{a-1}+n_i^a = N_i^a$ inclusive. Thus, the design consists of observations

$$Y_{ij}, i = 1, \ldots, m; j = N_i^{a-1} + 1, \ldots, N_i^{a-1} + n_i^a$$

for phases $a = 1, \ldots, 2k$.

As in the case of the $(AB)^1$ design, we posit a stochastic model in which the Y_{ij} are normally distributed, the data series for each individual lacks any time trend, and within individual, the deviations from the mean at each point in time are weakly stationary with first-order autocorrelation ϕ. Specifically, the statistical model for the a^{th} phase is

$$Y_{ij} = \frac{1}{2}\left[1 + (-1)^{a-1}\right]\mu^C + \frac{1}{2}[1 + (-1)^a]\mu^T + \eta_i + \varepsilon_{ij}$$

for $i = 1, \ldots, m$ and $j = N_i^{a-1}+1, \ldots, N_i^{a-1}+n_i^a$. The expressions in square brackets just assure that, in odd-numbered phases (baseline phases), the coefficient of μ^C is one and the coefficient of μ^T is zero and that in even-numbered phases (treatment phases), the coefficient of μ^T is one and the coefficient of μ^C is zero, where $\mu^T - \mu^C = \Delta$ represents the shift between baseline and treatment periods. We assume that the individual level effects η_i are normally distributed with zero mean and variance τ^2. The assumption that each time series is weakly stationary implies that, conditional on the η_i, the covariance of Y_{ij} with $Y_{i(j+t)}$ depends only on t. As before, we assume that the ε_{ij} have variance σ^2 and first-order autocorrelation ϕ within individuals, so that

$$\text{Cov}\left\{\varepsilon_{ij}, \varepsilon_{i'j'}\right\} = \begin{pmatrix} 0 & \text{if} \quad i \neq i' \\ \phi^{|j-j'|}\sigma^2 & \text{if} \quad i = i' \end{pmatrix}.$$

The denominator of the effect size estimate is based on the variance across individuals for each time point, averaged over time points. Because each individual can have a different number of observations within each phase, there may not be a complete set of m observations for some time points. The contribution to the variance for a time point is computed only if there is an observation at that time point for every individual. Define the minimum number of observations for any individual in the a^{th} phase by M^a, so that

$$M^a = \text{Minimum}\left\{n_1^a, \ldots, n_m^a\right\} \tag{16}$$

and define $M^{\bullet}$ to be the sum of the M^a, so that

$$M^{\bullet} = M^1 + \cdots + M^{2k}. \tag{17}$$

The variance pooled across phases and across individuals is

$$S^2 = \frac{1}{M^{\bullet}(m-1)} \sum_{a=1}^{2k} \sum_{j=N_i^{a-1}+1}^{N_i^{a-1}+M^a} \sum_{i=1}^{m} \left(Y_{ij} - \bar{Y}_{\bullet j}\right)^2 \tag{18}$$

where $\bar{Y}_{\bullet j}$ is the average across individuals of the j^{th} observations, given by

$$\bar{Y}_{\bullet j} = \frac{1}{m}\sum_{i=1}^{m} Y_{ij}.$$

Note that when $k = 1$, $n_i^a = n$ for $i = 1, \ldots, m$, and $a = 1,2$, it follows that $M^{\bullet} = 2n$ and expression (18) reduces to expression (4) for the pooled variance. The numerator of the effect size is the unweighted mean difference between phases A and B, defined as

$$\bar{D} = \frac{1}{mk}\sum_{i=1}^{m}\sum_{a=1}^{k}\left(\bar{Y}_{i\bullet}^{2a} - \bar{Y}_{i\bullet}^{2a-1}\right) = \frac{1}{mk}\sum_{i=1}^{m}\sum_{a=1}^{k}\left(\frac{1}{n_i^{2a}}\sum_{j=N_i^{2a-1}+1}^{N_i^{2a}} Y_{ij} - \frac{1}{n_i^{2a-1}}\sum_{j=N_i^{2a-2}+1}^{N_i^{2a-1}} Y_{ij}\right) \tag{19}$$

Define the effect size estimate ES exactly as in (1), namely,

$$\text{ES} = \frac{\bar{D}}{S} \tag{20}$$

where $\bar{D}$ is given in (19) and S^2 is given in (18).

Define the auxiliary constant A via

$$A = \frac{1}{k^2}\sum_{i=1}^{m}\sum_{a=1}^{2k}\sum_{b=1}^{2k}\left(\frac{(-1)^a(-1)^b}{n_i^a n_i^b}\sum_{s=N_i^{a-1}+1}^{N_i^a}\sum_{t=N_i^{b-1}+1}^{N_i^b}\phi^{|s-t|}\right) \tag{21}$$

and the auxiliary constants B, C, and D via

$$B = \sum_{i=1}^{m}\sum_{a=1}^{2k}\sum_{b=1}^{2k}\sum_{s=1}^{M^a}\sum_{t=1}^{M^b}\phi^{\left|N_i^{a-1}-N_i^{b-1}+s-t\right|} \tag{22}$$

$$C = \sum_{i=1}^{m}\sum_{a=1}^{2k}\sum_{b=1}^{2k}\sum_{s=1}^{M^a}\sum_{t=1}^{M^b}\phi^{2\left|N_i^{a-1}-N_i^{b-1}+s-t\right|} \tag{23}$$

and

$$D = \sum_{a=1}^{2k}\sum_{b=1}^{2k}\sum_{s=1}^{M^a}\sum_{t=1}^{M^b}\left(\sum_{i=1}^{m}\phi^{\left|N_i^{a-1}-N_i^{b-1}+s-t\right|}\right)^2. \tag{24}$$

The expected value of $\bar{D}$ is $\mu^T - \mu^C$, and the variance of $\bar{D}$ is

$$\mathrm{V}\{\bar{D}\} = \frac{A\sigma^2}{m^2}. \tag{25}$$

The expected value of S^2 is $\sigma^2 + \tau^2$, and the variance of S^2 is

$$\frac{2(\sigma^2+\tau^2)^2}{(M^\bullet)^2(m-1)}\left[(M^\bullet)^2\rho^2 + 2\rho(1-\rho)\left(\frac{B}{m}\right) + \frac{(1-\rho)^2}{m-1}\left\{\left(\frac{m-2}{m}\right)C + \frac{D}{m^2}\right\}\right] \tag{26}$$

where ρ is the intraclass correlation defined in (9). Proof of these facts is given in Appendix B.

If $\bar{D}$ and S^2 are independent, it follows by Box's [6] theorem that the sampling distribution of ES is a constant θ times a non-central t-distribution with v degrees of freedom, where θ is given by

$$\theta = \sqrt{\frac{\mathrm{V}\{\bar{D}\}}{\tau^2+\sigma^2}} = \frac{\sqrt{A(1-\rho)}}{m} \tag{27}$$

and v is given by

$$v = \frac{(M^\bullet)^2(m-1)^2}{(M^\bullet)^2(m-1)\rho^2 + 2\rho(1-\rho)\left(\frac{m-1}{m}\right)B + (1-\rho)^2\left[\left(\frac{m-2}{m}\right)C + \frac{D}{m^2}\right]}. \tag{28}$$

It follows from the results in Hedges [9] that the bias in ES can be corrected by multiplying ES by the correction factor J(v) defined in (11), so that the effect size

$$G = \mathrm{J}(v)\mathrm{ES} \tag{29}$$

is approximately an unbiased estimator of δ. It also follows that the variance of G is approximately

$$\mathrm{V}\{G\} = \mathrm{J}(v)^2\left[\frac{v\theta^2}{v-2} + \delta^2\left(\frac{v}{v-2} - \frac{1}{\mathrm{J}(v)^2}\right)\right]. \tag{30}$$

As a caveat, it should be noted that independence of $\bar{D}$ and S^2 might not hold in unbalanced $(\mathrm{AB})^k$ designs. Still, unless the degree of imbalance is severe, these approximations should remain fairly accurate.

Note that when $k = 1$ and $n_i^a = n$, for $i = 1, \ldots, m$; $a = 1,2$, so that we have an $(\mathrm{AB})^1$ design, the auxiliary constant $A = 2m(b_1 - c_1)$, where b_1 and c_1 are as previously defined. Also, note that under these restrictions, $B = 2mn^2(b_1 + c_1)$, $C = 2mn^2(b_2 + c_2)$, and $\mathrm{D} = 2m^2n^2(b_2 + c_2)$, where b_2 and c_2 are as previously defined. Thus, when both phases have equal numbers of observations for all individuals, the expressions for the effect size and its variance in the unbalanced $(\mathrm{AB})^k$ design reduce to the corresponding expressions given earlier for the $(\mathrm{AB})^1$ design.

6. Example 1

Here, we analyze data from an unbalanced $(\mathrm{AB})^2$ design reported by Lambert *et al.* [17]. This design has four phases: a baseline (control) phase followed by a treatment phase, another baseline phase, and another treatment

phase. Note that there are $m = 9$ cases, with each case having between 6 and 10 observations in each baseline phase and between 4 and 11 observations in each treatment phase. The minimum number of time points for the first phase is $M^1 = 6$, and the minimum numbers of time points for the second through fourth phases are $M^2 = 4$, $M^3 = 6$, and $M^4 = 7$, respectively, so that $M^\bullet = 23$. Approximately 10% of the data are missing, with missing observations scattered intermittently across cases and phases. For present purposes, missing data are ignored in the computations provided in the succeeding texts; a more thorough analysis of these data would use slight modifications to our methods to account for the missing data points. This example illustrates the computation of the effect size and its variance in an $(AB)^2$ design that is close to balanced, although not perfectly. The example was chosen because it has a relatively large number of replications across cases ($m = 9$).

The weighted average difference between phases is $\bar{D} = -5.458$, $S^2 = 4.674$, and $S = 2.162$. The estimate of effect size (before bias correction) is therefore

$$\text{ES} = \frac{-5.458}{2.162} = -2.525.$$

This estimate can be used to describe study results in the same metric as the standardized mean difference of a between-subjects study. It can also be combined with estimates from other studies in meta-analysis, both estimates from other single case designs and from between-subjects studies on the same question.

The results of this paper involving the autocorrelation structure can be used to calculate a correction for the estimation bias of ES as well as an estimate of effect size variance. Because the number of cases is relatively large for single case designs, the bias correction has little effect; however, this will not be true for many single case designs where the number of replications across cases is small. Following the method described in Appendix C, we estimate that the autocorrelation in these data is $\hat{\phi} = 0.225$. Following the method described in Appendix C, we obtain an estimate of the within-case variation $\hat{\sigma}^2 = 4.534$; combining this with $S^2 = 4.674$, we obtain an estimate of 0.030 for the intraclass correlation ρ. Using the autocorrelation $\phi = 0.225$, the values of the auxiliary constants are $A = 1.754$, $B = 294.751$, $C = 223.488$, and $D = 2002.444$. Inserting the value $\rho = 0.030$ and the values of the auxiliary constants B, C, and D into (28), we obtain $\nu = 164.492$ degrees of freedom. This value of the degrees of freedom permits us to compute G, the bias-corrected estimate of effect size. Using $\text{ES} = -2.525$ and inserting 164.492 degrees of freedom into (29), we obtain $G = -2.513$. Inserting the value of the auxiliary constant A and $\rho = 0.030$ into (27), we obtain the value $\theta = 0.145$. Finally, inserting the values ν, θ, and $\delta = G = -2.513$ into (30), we obtain $\text{V}\{G\} = 0.041$.

Because the estimation of the variance and bias correction involves estimation of the nuisance parameters ϕ and ρ, it is useful to see how plausible variation in the estimates of these parameters might affect the bias correction and the variance estimates. The large-sample variance of the maximum likelihood estimator of ϕ is 0.004, which corresponds to a standard error of 0.060; a range of plausible values (a range of two standard errors around the estimate) for ϕ is therefore about 0.10–0.35. Holding fixed the values of ES, S^2, and the within-case, within-phase sample variance $\hat{\gamma}_\bullet^\bullet(0)$ (defined in Appendix C), we now examine how varying ϕ affects our estimates of ν, G, and $\text{V}\{G\}$. The estimated degrees of freedom ν ranges from 152.6 to 164.5. This variation has only a trivial impact on the effect size estimate: G ranges only from -2.512 to -2.513. The impact on the variance estimate is also minor, with $\text{V}\{G\}$ ranging from 0.037 to 0.047.

7. Example 2

Here, we analyze data from an unbalanced $(AB)^2$ design reported in Anglesea *et al*. [3]. This design has four phases: a baseline (control) phase followed by a treatment phase, another baseline phase, and a final treatment phase. Note that there are $m = 3$ cases with $n_1^1 = 7$, $n_1^2 = 6$, $n_1^3 = 7$, $n_1^4 = 7$ for the first case; $n_2^1 = 4$, $n_2^2 = 4$, $n_2^3 = 3$, $n_2^4 = 3$ for the second case; and $n_3^1 = 4$, $n_3^2 = 4$, $n_3^3 = 4$, $n_2^4 = 2$ for the third case. Here, the minimum number of time points for the first phase is $M^1 = 4$; the minimum number of time points for the second phase is $M^2 = 4$; the minimum number of time points for the third phase is $M^3 = 3$; and the minimum number of time points for the fourth phase is $M^4 = 2$; so that $M^\bullet = 13$. This example was chosen because it is extreme, with just enough replications across cases ($m = 3$) to permit estimation of a variance.

The weighted average difference between phases is $\bar{D} = 86.870$, $S^2 = 2347.8$, and $S = 48.455$. The estimate of effect size (before bias correction) is therefore

$$\text{ES} = \frac{86.870}{48.455} = 1.793.$$

This estimate can be used to describe study results in the same metric as a standardized mean difference from a between-subjects study and can be combined with estimates from other studies in meta-analysis.

Although using a bias-corrected effect size typically has little effect in between-subjects studies, this is not true for the present example. The estimated autocorrelation in these data is again small: the corrected Yule–Walker estimate is $\hat{\phi} = 0.176$. The estimated within-cases variation is $\hat{\sigma}^2 = 198.4$; using this estimate and S^2, we obtain

an estimate of 0.916 for ρ. Using the autocorrelation $\phi = 0.176$, the values of the auxiliary constants are $A = 0.889$, $B = 52.162$, $C = 41.030$, and $D = 122.725$. Inserting $\rho = 0.916$ and the values of the auxiliary constants B, C, and D into (28), we obtain $\nu = 2.340$ degrees of freedom. Using ES = 1.793 and inserting 2.340 degrees of freedom into (29), we obtain $G = 1.150$. Inserting the value of the auxiliary constant A and $\rho = 0.916$ into (27), we obtain the value $\theta = 0.091$. Finally, inserting the values ν, θ, and $\delta = G = 1.150$ into (30), we obtain $V\{G\} = 2.440$.

Because estimation of the bias correction and variance involves the nuisance parameters, it is useful to see how plausible variation in the estimates of the nuisance parameters might affect the bias correction and the variance estimates. The large sample variance of the maximum likelihood estimator of ϕ is 0.019, which corresponds to a standard error of 0.137; a range of plausible values for ϕ is therefore about -0.10 to 0.45. Using $\phi = -0.10$, we obtain estimates of $\nu = 2.310$ degrees of freedom and a bias-corrected effect size of $G = 1.140$, with variance estimate $V\{G\} = 2.636$. In comparison, using $\phi = 0.45$ leads to $\nu = 2.399$ degrees of freedom, $G = 1.167$, and $V\{G\} = 2.140$. The impact of plausible variation of the nuisance parameters on the estimate (G) is small, but the impact on the variance is not. This is largely because the degrees of freedom are so small; one or two additional degrees of freedom would have substantially reduced the influence of the nuisance parameters on the variance. It is generally advisable to have at least a few additional case beyond $m = 3$.

8. Conclusion

We have introduced an effect size measure for one kind of effect in single case designs (mean shift between baseline and treatment phases) that estimates the same parameter as in the corresponding between-subjects design. The estimator itself depends on nuisance parameters only for a bias correction term. Even when nuisance parameters are estimated rather crudely, the estimator has relatively small bias. This makes the proposed estimation techniques suitable for use even in designs with as few as three independent cases, as in the second example earlier. However, the proposed variance estimator tends to have somewhat larger bias, typically being underestimated when nuisance parameters are estimated.

The problem of effect size estimates whose variances depend on nuisance parameters is not unknown in meta-analysis. One example is when synthetic composite effect size estimates are created to 'average' or 'difference' correlated estimates within studies (see, e.g. [11]). The variance of the composite depends on the correlation structure of the estimates, which is typically unknown in detail. In this case, meta-analysts usually construct a variance estimate for the composite on the basis of the values of the correlations chosen so that the estimate is conservative (i.e. overestimating the variance of the composite effect size). Another example is when effect sizes are adjusted for clustering in experiments that involve multilevel sampling (see, e.g. [10]). In this case, meta-analysts usually choose values of the intraclass correlation that are designed to be conservative, yielding overestimates of the variance of the effect size. Choosing conventional but conservative values of the nuisance parameters in this case would be consistent with these precedents.

If the estimates proposed here are used in meta-analyses, it would be advisable to use at least partially empirical variance estimation procedures, such as random effects models. Entirely empirical variance estimates for meta-analysis, such as those based on bootstrapping or randomization tests (e.g. [1]) or fully empirical robust standard errors [13], could also be used in the meta-analysis. If this approach is taken, the analyst can avoid entirely computations of variances of individual effect size estimates (except perhaps for crude decisions about weighting). If effect size estimates based on single case designs are included in meta-analyses alongside those from between-subjects designs, the latter are likely to receive much higher weight given their typically much larger sample sizes. Small biases in the variance of estimates from single case designs may therefore have relatively little impact on the overall analysis.

Our statistical model and estimation procedures assume that the series shows no trend over time. Although this assumption appears plausible in many single case designs that we have seen, [23] little research has investigated the presence and form of trend empirically. If the researcher doubts the validity of this assumption in a particular application, there are several options. The easiest option to execute is to de-trend the data with regression, then analyze the residuals. However, de-trending (or other techniques such as taking first differences) can implicitly change the underlying modeling assumptions. Our preferred approach would be to make such assumptions explicit, by directly specifying a trend component in the statistical model. We anticipate that the methods described in this paper can be extended to a model with a time trend that is common across both phases and cases. However, more elaborate modeling assumptions, such as time trends that differ across phases or vary randomly across cases, may require specifying new and further effect size parameters.

Similarly, our model assumes a first-order autocorrelation process. Again, little research exists about whether such an assumption holds, or whether higher order autocorrelations might be present in single case design data. However, the first-order assumption is common in this kind of research, exactly because of lack of either evidence or strong theory leading us to expect the contrary. This is a fruitful area for future research.

Future research may also provide improved estimators of the variance of the effect size estimates proposed in this paper, as well as of related estimators. For example, our preliminary work suggests that bias corrections may

substantially improve estimation of autocorrelations when the total number of observations for each individual is small. The use of external information about autocorrelations to stabilize estimates (e.g. via empirical Bayes estimation) also has potential to improve variance estimation.

Finally, the fact that the standard errors of study-level effect sizes can be large, especially when the number of cases is minimal, means that study effects may be statistically non-significant. This may dismay researchers, especially those who are used to judging effect size visually even when the number of case is small. We can say four things to ameliorate such concerns. First, our method is not intended to replace visual analysis but to complement it, while at the same time providing some statistical sense of how much confidence in the effect should be influenced by sampling error. Second, hopefully, both the present work and emerging standards for single case designs (e.g. [15]) will encourage single case design researchers to see the benefits of increasing the number of cases in each study. Even relatively small increases are likely to ameliorate the power problem greatly. Third, increasing the number of data points within each case will also increase power and may be especially helpful when increasing the number of cases is not feasible. Finally, the present method also allows single case researchers to aggregate the results of multiple studies on the same question, even when the dependent variables are in different metrics, using meta-analytic techniques. The meta-analysis of studies using single case designs, whether combined with results from between-subjects experiments or not, will greatly increase the power of the resulting effect size estimates (e.g. [21]).

Appendix A

The exact expectation and variance of ES can be derived from the moments of $\bar{D}$ and the inverse moments of S^2. To simplify the notation, we limit this presentation to the balanced $(AB)^1$ design. First, observe that S^2 is a quadratic form of the $2nm$-dimensional normal random variable $\mathbf{y}$ with covariance matrix $\mathbf{\Sigma}$, so that $2n(m-1)S^2 = \mathbf{y}'\mathbf{A}\mathbf{y} = \mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y}$. Then, it follows that $\mathbf{A}\mathbf{y} \sim N(\mathbf{0}, \mathbf{A}\mathbf{\Sigma}\mathbf{A})$. The covariance matrix $\mathbf{A}\mathbf{\Sigma}\mathbf{A}$ has $2n$ unique nonzero eigenvalues $\lambda_1, \ldots, \lambda_{2n}$, each repeated $m-1$ times, and each a function of ϕ, ρ, m, and n. It follows by Theorem 3.2b.4 in Mathai and Provost [19] that, for values $h < n(m-1)$,

$$E\left\{\left(\mathbf{y}'\mathbf{A}\mathbf{y}\right)^{-h}\right\} = \frac{\alpha^h}{\Gamma(h)}\int_0^1 u^{h-1}(1-u)^{n(m-1)-h-1}\prod_{j=1}^{2n}\left[1-u\left(1-2\alpha\lambda_j\right)\right]^{-\frac{m-1}{2}}du$$

where α is an arbitrary constant such that $|1-2\alpha\lambda_j| < 1$ for $j = 1, \ldots, m-1$. This expression can be evaluated numerically. Because $\bar{D}$ is independent of S, the moments of ES are the products of the moments of $\bar{D}$ and the inverse moments of S.

Appendix B

Here, we give the derivation of the effect size estimator for the general $(AB)^k$ design, allowing unequal number of time points per phase and per individual. The expected value and variance of $\bar{D}$ follow from basic properties of the multivariate normal distribution. It remains to find the sampling distribution of S^2.

Partition the data $mM^\bullet \times 1$ vector of observations into $2k$ subvectors so that

$$\mathbf{y}' = \left(\mathbf{y}_\bullet^{1'}, \ldots, \mathbf{y}_\bullet^{2k'}\right)$$

where the vector $\mathbf{y}_\bullet^a$ is the vector of mM^a observations in the a^{th} phase. Further, partition $\mathbf{y}_\bullet^a$ into the observations for the M^a time points (shared across all individuals) in the a^{th} phase as

$$\mathbf{y}_\bullet^{a'} = \left(\mathbf{y}_1^{a'}, \ldots, \mathbf{y}_{M^a}^{a'}\right), a = 1, \ldots, 2k$$

where the vector $\mathbf{y}_s^a$ of m observations at the s^{th} time point is given by

$$\mathbf{y}_s^{a'} = \left(Y_{1\left(N_1^{a-1}+s\right)}, \ldots, Y_{m\left(N_m^{a-1}+s\right)}\right), s = 1, \ldots, M^a.$$

The complexity of this partition arises for two reasons. First, the number of time points per phase shared across all individuals, M^a, need not be the same across phases (if it were, we could set $M^a = M$). Secondly, the number of time points per individual in each phase is not the same across either individuals or across phases. Moreover, the number of time points per individual need not be the same as the number of time points shared across individuals in any phase. For example, the first time point in a phase may be the fifth time point overall for individual i but the eighth time point for individual j.

Partition the $mM^{\bullet} \times mM^{\bullet}$ covariance matrix $\mathbf{\Sigma}_{\mathrm{T}}$ of $\mathbf{y}$ into a $2k \times 2k$ partitioned matrix

$$\mathbf{\Sigma}_{\mathrm{T}} = \left(\begin{array}{c|c|c} \mathbf{\Sigma}_{\mathrm{T}}^{11} & \cdots & \mathbf{\Sigma}_{\mathrm{T}}^{1(2k)} \\ \hline \vdots & \ddots & \vdots \\ \hline \mathbf{\Sigma}_{\mathrm{T}}^{(2k)1} & \cdots & \mathbf{\Sigma}_{\mathrm{T}}^{(2k)(2k)} \end{array}\right)$$

where $\mathbf{\Sigma}_{\mathrm{T}}^{ab}$ is the $mM^{a} \times mM^{b}$ matrix of covariances of $\mathbf{y}_{\bullet}^{a}$ with $\mathbf{y}_{\bullet}^{b}$. Further, partition each matrix $\mathbf{\Sigma}_{\mathrm{T}}^{ab}$ into a total of $M^{a}M^{b}$ submatrices each of dimension $m \times m$ as follows

$$\mathbf{\Sigma}_{\mathrm{T}}^{ab} = \left(\begin{array}{c|c|c} \mathbf{\Sigma}_{11}^{ab} & \cdots & \mathbf{\Sigma}_{1M^{b}}^{ab} \\ \hline \vdots & \ddots & \vdots \\ \hline \mathbf{\Sigma}_{M^{a}1}^{ab} & \cdots & \mathbf{\Sigma}_{M^{a}M^{b}}^{ab} \end{array}\right)$$

where $\mathbf{\Sigma}_{st}^{ab}$ is the covariance matrix of $\mathbf{y}_{s}^{a}$ with $\mathbf{y}_{t}^{b}$. Because observations from different individuals are independent, each of the $\mathbf{\Sigma}_{st}^{ab}$ is an $m \times m$ diagonal matrix. That is, the covariances between any observation in phase a for individual i and any observation in phase b for individual j is zero if $i \neq j$. Only the covariances between different observations on the same individual are non-zero. However, because of the imbalance in the design, the diagonal elements of $\mathbf{\Sigma}_{st}^{ab}$ are not equal. This is because the covariance between observations on the same individual depends on the number of observations separating them, but the s^{th} observation in phase a for individual i (observation $N_i^{a-1} + s$) is separated from the t^{th} observations in phase b for individual i (observation $N_i^{b-1} + t$) by $(N_i^{a-1} + s - N_i^{b-1} - t)$ observations. Because this quantity depends on i, it is different for each individual.

The matrix $\mathbf{\Sigma}_{st}^{ab}$ can be written as the sum of two diagonal matrices

$$\Sigma_{st}^{ab} = \tau^2 \mathbf{I}_m + \sigma^2 \mathbf{D}_{st}{}^{ab},$$

where the the i^{th} diagonal element of $\mathbf{D}_{st}^{ab}$ is the autocorrelation ϕ raised to the power $d_{iabst} = |N_i^{a-1} - N_i^{b-1} + s - t|$, that is,

$$\mathbf{D}_{st}^{ab} = \operatorname{diag}\left(\varphi^{d_{1abst}}, \ldots, \varphi^{d_{mabst}}\right).$$

Note that $\mathbf{D}_{st}^{ab} = \mathbf{D}_{ts}^{ba}$, because interchanging both a and b and s and t simply changes the sign of the exponent within the absolute value.

The quadratic form S^2 can be written as $S^2 = \mathbf{y}'\mathbf{A}_{\mathrm{T}}\mathbf{y}/M^{\bullet}(m-1)$, where the matrix $\mathbf{A}_{\mathrm{T}}$ can be partitioned conformably to $\mathbf{\Sigma}_{\mathrm{T}}$, that is, partition the $mM^{\bullet} \times mM^{\bullet}$ matrix $\mathbf{A}_{\mathrm{T}}$ into a $2k \times 2k$ block diagonal partitioned matrix

$$\mathbf{A}_{\mathrm{T}} = \left(\begin{array}{ccc} \mathbf{A}^{1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{A}^{2k} \end{array}\right)$$

where $\mathbf{A}^{a}$ is the $mM^{a} \times mM^{a}$ block diagonal matrix. Further, partition each matrix $\mathbf{A}^{a}$ into a total of $(M^{a})^2$ submatrices each of dimension $m \times m$ as follows

$$\mathbf{A}^{a} = \left(\begin{array}{c|c|c} \mathbf{A}_{1} & \cdots & 0 \\ \hline \vdots & \ddots & \vdots \\ \hline 0 & \cdots & \mathbf{A}_{M^{a}} \end{array}\right)$$

where $\mathbf{A}_s$ is an $m \times m$ matrix defined by $\mathbf{A}_s = \mathbf{A} = \mathbf{I}_m - \mathbf{1}_m\mathbf{1}_m'/m$ and $\mathbf{I}_m$ is an $m \times m$ identity matrix and $\mathbf{1}_m$ is a $m \times 1$ vector of 1's.

The expected value of $M^{\bullet}(m-1)S^2$, the numerator of S^2, is

$$\mathrm{E}\left\{\left(M^{\bullet}\left(m-1\right)\right)S^{2}\right\} = \operatorname{tr}\left(\mathbf{A}_{T}\Sigma_{\mathrm{T}}\right) = \operatorname{tr}\left(\begin{array}{c|c|c} \mathbf{A}^{1}\mathbf{\Sigma}_{\mathrm{T}}^{11} & \cdots & \mathbf{A}^{1}\mathbf{\Sigma}_{\mathrm{T}}^{1(2k)} \\ \hline \vdots & \ddots & \vdots \\ \hline \mathbf{A}^{2k}\mathbf{\Sigma}_{\mathrm{T}}^{(2k)1} & \cdots & \mathbf{A}^{2k}\mathbf{\Sigma}_{\mathrm{T}}^{(2k)(2k)} \end{array}\right).$$

Now, each $\mathbf{A}^{a}\mathbf{\Sigma}_{\mathrm{T}}^{ab}$ is of form

$$\mathbf{A}^{a}\mathbf{\Sigma}_{\mathrm{T}}^{ab} = \left(\begin{array}{c|c|c} \mathbf{A}\mathbf{\Sigma}_{11}^{ab} & \cdots & \mathbf{A}\mathbf{\Sigma}_{1M^{b}}^{ab} \\ \hline \vdots & \ddots & \vdots \\ \hline \mathbf{A}\mathbf{\Sigma}_{M^{a}1}^{ab} & \cdots & \mathbf{A}\mathbf{\Sigma}_{M^{a}M^{b}}^{ab} \end{array}\right).$$

Therefore, the expected value of $M^{\bullet}(m-1)S^2$ is

$$\sum_{a=1}^{2k}\mathrm{tr}\left(\mathbf{A}^{a}\boldsymbol{\Sigma}_{\mathrm{T}}^{aa}\right)=\sum_{a=1}^{2k}\sum_{s=1}^{M^{a}}\mathrm{tr}\left(\mathbf{A}\boldsymbol{\Sigma}_{ss}^{aa}\right)=\sum_{a=1}^{2k}\sum_{s=1}^{M^{a}}\left(\sigma^{2}+\tau^{2}\right)(m-1)=M^{\bullet}(m-1)\left(\sigma^{2}+\tau^{2}\right)$$

and the expected value of $S^2=\sigma^2+\tau^2$.

The variance of S^2 can be obtained from the variance of $M^{\bullet}(m-1)S^2$, the numerator of S^2, as

$$\mathrm{V}\left(M^{\bullet}(m-1)S^{2}\right)=2\mathrm{tr}\left(\mathbf{A}_{T}\boldsymbol{\Sigma}_{\mathrm{T}}\mathbf{A}_{T}\boldsymbol{\Sigma}_{\mathrm{T}}\right)$$

$$=2\mathrm{tr}\left\{\left(\begin{array}{c|c|c}\mathbf{A}^{1}\boldsymbol{\Sigma}_{\mathrm{T}}^{11} & \cdots & \mathbf{A}^{1}\boldsymbol{\Sigma}_{\mathrm{T}}^{1(2k)}\\ \hline \vdots & \ddots & \vdots\\ \hline \mathbf{A}^{2k}\boldsymbol{\Sigma}_{\mathrm{T}}^{(2k)1} & \cdots & \mathbf{A}^{2k}\boldsymbol{\Sigma}_{\mathrm{T}}^{(2k)(2k)}\end{array}\right)\left(\begin{array}{c|c|c}\mathbf{A}^{1}\boldsymbol{\Sigma}_{\mathrm{T}}^{11} & \cdots & \mathbf{A}^{1}\boldsymbol{\Sigma}_{\mathrm{T}}^{1(2k)}\\ \hline \vdots & \ddots & \vdots\\ \hline \mathbf{A}^{2k}\boldsymbol{\Sigma}_{\mathrm{T}}^{(2k)1} & \cdots & \mathbf{A}^{2k}\boldsymbol{\Sigma}_{\mathrm{T}}^{(2k)(2k)}\end{array}\right)\right\}$$

$$=2\mathrm{tr}\left(\begin{array}{c|c|c}\displaystyle\sum_{a=1}^{2k}\mathbf{A}^{1}\boldsymbol{\Sigma}_{\mathrm{T}}^{1a}\mathbf{A}^{a}\boldsymbol{\Sigma}_{\mathrm{T}}^{a1} & \cdots & \displaystyle\sum_{a=1}^{2k}\mathbf{A}^{1}\boldsymbol{\Sigma}_{\mathrm{T}}^{1a}\mathbf{A}^{a}\boldsymbol{\Sigma}_{\mathrm{T}}^{a(2k)}\\ \hline \vdots & \ddots & \vdots\\ \hline \displaystyle\sum_{a=1}^{2k}\mathbf{A}^{2k}\boldsymbol{\Sigma}_{\mathrm{T}}^{(2k)a}\mathbf{A}^{a}\boldsymbol{\Sigma}_{\mathrm{T}}^{a1} & \cdots & \displaystyle\sum_{a=1}^{2k}\mathbf{A}^{(2k)}\boldsymbol{\Sigma}_{\mathrm{T}}^{(2k)a}\mathbf{A}^{a}\boldsymbol{\Sigma}_{\mathrm{T}}^{a(2k)}\end{array}\right)$$

so that

$$2\mathrm{tr}\left(\mathbf{A}_{T}\boldsymbol{\Sigma}_{\mathrm{T}}\mathbf{A}_{T}\boldsymbol{\Sigma}_{\mathrm{T}}\right)=2\sum_{a=1}^{2k}\sum_{b=1}^{2k}\mathrm{tr}\left(\mathbf{A}^{b}\boldsymbol{\Sigma}_{\mathrm{T}}^{ba}\mathbf{A}^{a}\boldsymbol{\Sigma}_{\mathrm{T}}^{ab}\right).$$

Now, each $\mathbf{A}^{b}\boldsymbol{\Sigma}_{\mathrm{T}}{}^{ba}\mathbf{A}^{a}\boldsymbol{\Sigma}_{\mathrm{T}}{}^{ab}$ is of form

$$\mathbf{A}^{b}\boldsymbol{\Sigma}_{\mathrm{T}}^{ba}\mathbf{A}^{a}\boldsymbol{\Sigma}_{\mathrm{T}}^{ab}=\left(\begin{array}{c|c|c}\mathbf{A}\boldsymbol{\Sigma}_{11}^{ba} & \cdots & \mathbf{A}\boldsymbol{\Sigma}_{1M^{a}}^{ba}\\ \hline \vdots & \ddots & \vdots\\ \hline \mathbf{A}\boldsymbol{\Sigma}_{M^{b}1}^{ba} & \cdots & \mathbf{A}\boldsymbol{\Sigma}_{M^{b}M^{a}}^{ba}\end{array}\right)\left(\begin{array}{c|c|c}\mathbf{A}\boldsymbol{\Sigma}_{11}^{ab} & \cdots & \mathbf{A}\boldsymbol{\Sigma}_{1M^{b}}^{ab}\\ \hline \vdots & \ddots & \vdots\\ \hline \mathbf{A}\boldsymbol{\Sigma}_{M^{a}1}^{ab} & \cdots & \mathbf{A}\boldsymbol{\Sigma}_{M^{a}M^{b}}^{ab}\end{array}\right)$$

$$=\left(\begin{array}{c|c|c}\displaystyle\sum_{s=1}^{M^{a}}\mathbf{A}\boldsymbol{\Sigma}_{1s}^{ba}\mathbf{A}\boldsymbol{\Sigma}_{s1}^{ab} & \cdots & \displaystyle\sum_{s=1}^{M^{a}}\mathbf{A}\boldsymbol{\Sigma}_{1s}^{ba}\mathbf{A}\boldsymbol{\Sigma}_{sM^{b}}^{ab}\\ \hline \vdots & \ddots & \vdots\\ \hline \displaystyle\sum_{s=1}^{M^{a}}\mathbf{A}\boldsymbol{\Sigma}_{M^{b}s}^{ba}\mathbf{A}\boldsymbol{\Sigma}_{s1}^{ab} & \cdots & \displaystyle\sum_{s=1}^{M^{a}}\mathbf{A}\boldsymbol{\Sigma}_{M^{b}s}^{ba}\mathbf{A}\boldsymbol{\Sigma}_{sM^{b}}^{ab}\end{array}\right).$$

Therefore,

$$2\mathrm{tr}\left(\mathbf{A}_{T}\boldsymbol{\Sigma}_{\mathrm{T}}\mathbf{A}_{T}\boldsymbol{\Sigma}_{\mathrm{T}}\right)=2\sum_{a=1}^{2k}\sum_{b=1}^{2k}\sum_{s=1}^{M^{a}}\sum_{t=1}^{M^{b}}\mathrm{tr}\left(\mathbf{A}\boldsymbol{\Sigma}_{ts}^{ba}\mathbf{A}\boldsymbol{\Sigma}_{st}^{ab}\right).$$

Recall that $\boldsymbol{\Sigma}_{st}^{ab}=\tau^{2}\mathbf{I}_{m}+\sigma^{2}\mathbf{D}_{st}^{ab}$, so that

$$\mathbf{A}\boldsymbol{\Sigma}_{ts}^{ba}\mathbf{A}\boldsymbol{\Sigma}_{st}^{ab}=\mathbf{A}\left(\tau^{2}\mathbf{I}+\sigma^{2}\mathbf{D}_{ts}^{ba}\right)\mathbf{A}\left(\tau^{2}\mathbf{I}+\sigma^{2}\mathbf{D}_{st}^{ab}\right)=\tau^{4}\mathbf{A}^{2}+\tau^{2}\sigma^{2}\mathbf{A}^{2}\mathbf{D}_{st}^{ab}+\tau^{2}\sigma^{2}\mathbf{A}\mathbf{D}_{ts}^{ba}\mathbf{A}+\sigma^{4}\mathbf{A}\mathbf{D}_{ts}^{ba}\mathbf{A}\mathbf{D}_{st}^{ab}.$$

Using the facts that $\mathbf{A}$ is idempotent and that $\mathbf{D}_{st}^{ab}=\mathbf{D}_{ts}^{ba}$ is diagonal, we see that

$$\mathrm{tr}\left(\mathbf{A}\boldsymbol{\Sigma}_{ts}^{ba}\mathbf{A}\boldsymbol{\Sigma}_{st}^{ab}\right)=(m-1)\tau^{4}+2\sigma^{2}\tau^{2}\left(\frac{m-1}{m}\right)\mathrm{tr}\left(\mathbf{D}_{st}^{ab}\right)+\sigma^{4}\mathrm{tr}\left(\mathbf{A}\mathbf{D}_{ts}^{ba}\mathbf{A}\mathbf{D}_{st}^{ab}\right).$$

Recalling that $\mathbf{A}=\mathbf{I}_{m}-\mathbf{1}_{m}\mathbf{1}_{m}'/m$ and using the elementary properties of the trace, we obtain an expression for the trace of $\mathbf{A}\mathbf{D}_{ts}^{ab}\mathbf{A}\mathbf{D}_{st}^{ab}$ in the last term as

$$\mathrm{tr}\left(\mathbf{A}\mathbf{D}_{ts}^{ba}\mathbf{A}\mathbf{D}_{st}^{ab}\right)=\mathrm{tr}\left(\left[\mathbf{D}_{st}^{ab}\right]^{2}\right)-\frac{2}{m}\mathrm{tr}\left(\left[\mathbf{D}_{st}^{ab}\right]^{2}\right)+\frac{1}{m^{2}}\mathrm{tr}\left(\mathbf{1}_{m}\mathbf{1}_{m}{}'\mathbf{D}_{ts}^{ba}\mathbf{1}_{m}\mathbf{1}_{m}{}'\mathbf{D}_{st}^{ab}\right)=\left(\frac{m-2}{m}\right)\mathrm{tr}\left(\left[\mathbf{D}_{st}^{ab}\right]^{2}\right)+\frac{1}{m^{2}}\left[\mathrm{tr}\left(\mathbf{D}_{st}^{ab}\right)\right]^{2}$$

Using the fact that $\mathbf{D}_{ts}^{ab}$ is diagonal, we see that

$$\mathrm{tr}\left(\mathbf{D}_{st}^{ab}\right) = \sum_{i=1}^{m} \phi^{d_{iabst}}$$

and

$$\mathrm{tr}\left(\left[\mathbf{D}_{st}^{ab}\right]^2\right) = \sum_{i=1}^{m} \phi^{2d_{iabst}}$$

so that

$$\mathrm{tr}\left(\mathbf{A}\boldsymbol{\Sigma}_{ts}^{ba}\mathbf{A}\boldsymbol{\Sigma}_{st}^{ab}\right) = (m-1)\tau^4 + 2\sigma^2\tau^2\left(\frac{m-1}{m}\right)\sum_{i=1}^{m}\phi^{d_{iabst}} + \sigma^4\left\{\left(\frac{m-2}{m}\right)\sum_{i=1}^{m}\phi^{2d_{iabst}} + \frac{1}{m^2}\left[\sum_{i=1}^{m}\phi^{d_{iabst}}\right]^2\right\}.$$

Summing this expression over $a = 1,\ldots,2k$; $b = 1,\ldots,2k$; $s = 1,\ldots,M^a$; and $t = 1,\ldots,M^b$, then rearranging terms, we obtain expression (26).

Appendix C

Here, we give formulas for estimators of ϕ and ρ in the unbalanced $(AB)^k$ design, of which the $(AB)^1$ design is a special case. Define the sample auto-covariance for case i and phase a as a function of the lag h:

$$\hat{\gamma}_i^a(h) = \frac{1}{n_i^a}\sum_{j=N_i^{a-1}+1}^{N_i^a-h}\left(Y_{ij} - \bar{Y}_{i\bullet}^a\right)\left(Y_{i(j+h)} - \bar{Y}_{i\bullet}^a\right)$$

where $\bar{Y}_{i\bullet}^a$ is the average across observations of the i^{th} individual within the a^{th} phase given by

$$\bar{Y}_{i\bullet}^a = \frac{1}{n_i^a}\sum_{j=N_i^{a-1}+1}^{N_i^a} Y_{ij}.$$

For improved precision, these can be pooled across phases and cases, yielding

$$\hat{\gamma}_{\bullet}^{\bullet}(h) = \frac{1}{N_{\bullet}^{\bullet}}\sum_{i=1}^{m}\sum_{a=1}^{2k} n_i^a\hat{\gamma}_i a(h).$$

An estimate of ϕ is then given by

$$\hat{\phi} = \frac{\hat{\gamma}_{\bullet}^{\bullet}(1)}{\hat{\gamma}_{\bullet}^{\bullet}(0)} + c$$

where the constant c is given by

$$c = \frac{\sum_{i=1}^{m}\sum_{a=1}^{2k}\left(1 - \frac{1}{n_i^a}\right)}{\sum_{i=1}^{m}\sum_{a=1}^{2k}\left(n_i^a - 1\right)} = \frac{2km - \sum_{i=1}^{m}\sum_{a=1}^{2k} 1/n_i^a}{N_{\bullet}^{\bullet} - 2km}.$$

The use of the c correction makes $\hat{\phi}$ approximately unbiased when $\phi = 0$. However, for non-null ϕ, the estimate remains biased towards zero, particularly so when each phase is short. Note that for balanced designs (in which $n_i^a = n$ for $i = 1,\ldots,m$ and $a = 1,\ldots,2k$), the constant simplifies to $c = 1/n$, the correction studied by Huitema and McKean [14].

An estimate of σ^2 is needed for the purpose of estimating ρ. The Yule–Walker estimate of σ^2 is simply $\hat{\gamma}_{\bullet}^{\bullet}(0)$, the within-phase-within-individual variance (using divisor $N_{\bullet}^{\bullet}$ rather than $N_{\bullet}^{\bullet} - 2mk$). The expected value of $\hat{\gamma}_{\bullet}^{\bullet}(0)$ is $E\sigma^2/N_{\bullet}^{\bullet}$, where

$$E = \sum_{i=1}^{m}\sum_{a=1}^{2k}\left(n_i^a - \frac{1}{n_i^a}\sum_{s=1}^{n_i^a}\sum_{t=1}^{n_i^a}\phi^{|s-t|}\right).$$

It follows that $\hat{\sigma}^2 = N_{\bullet}^{\bullet}\gamma_{\bullet}^{\bullet}(0)/E$ is an unbiased estimator of σ^2 for known ϕ. Note that in the balanced $(AB)^1$ design, $E/N_{\bullet}^{\bullet}$ reduces to $(1 - b_1)$. We then estimate ρ via

$$\hat{\rho} = 1 - \frac{\hat{\sigma}^2}{S^2}$$

Acknowledgements

This research was supported in part by the Institute of Education Sciences, U.S. Department of Education, through grants R305D100046 to the University of California, Merced and R305B080027 to Northwestern

University. The opinions expressed are those of the author and do not represent views of the Institute or the US Department of Education.

References

1. Adams DC, Gurevitch J, Rosenberg MS. 1997. Resampling tests for meta-analysis of ecological data. *Ecology* **78**: 1277–1283.
2. Allison DB, Gorman BS. 1993. Calculating effect sizes for meta-analysis: The case of the single case. *Behav Res Ther* **31**: 621–631.
3. Anglesea MM, Hooch H, Taylor BA. 2008. Reducing rapid eating in teenagers with autism: Use of a page prompt. *J Appl Behav Anal* **41**: 107–111.
4. Beretvas SN, Chung H. 2008. A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention* **2**: 129–141.
5. Borenstein M. 2009. Effect sizes for continuous data, in *The Handbook of Research Synthesis and Meta-Analysis* Cooper H, Hedges LV, Valentine J. (eds). Russell Sage Foundation, New York, pp 221–235.
6. Box GEP. 1954. Some theorems on quadratic forms applied to the study of analysis of variance problems, I: Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics* **25**: 290–302.
7. Cooper H, Hedges LV, Valentine J. 2009. *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation, New York.
8. Fleiss JL, Berlin JA. 2009. Effect sizes for dichotomous data, in *The Handbook of Research Synthesis and Meta-Analysis* Cooper H, Hedges LV, Valentine J (eds). Russell Sage Foundation, New York, pp 238–253.
9. Hedges LV. 1981. Distribution theory for Glass's estimator of effect size and related estimators. *J Educational Statistics* **6**: 107–128.
10. Hedges LV. 2007a. Effect sizes in cluster randomized designs. *J Educational and Behavioral Statistics* **32**: 341–370.
11. Hedges LV. 2007b. Meta-analysis, in *The Handbook of Statistics, Vol. 26, Psychometrics* Rao CR, Sinharay S (eds). Amsterdam, pp 919–953. Elsevier.
12. Hedges LV. 2008. What are effect sizes and why do we need them? *Developmental Psychology Perspectives* **2**: 167–171.
13. Hedges LV, Tipton E, Johnson M. 2010. Robust variance estimation for meta-regression with dependent effect size estimators. *J Research Synthesis Methods* **1**: 39–65.
14. Huitema BE, McKean JW. 1991. Autocorrelation estimation and inference with small samples. *Psychol Bull* **110**: 291–304.
15. Kratochwill TR, Hitchcock J, Horner RH, Levin JR, Odom SL, Rindskopf DM, Shadish WR. 2010. Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
16. Kratochwill TR, Levin JR. 2010. Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychol Methods* **15**: 124–144.
17. Lambert MC, Cartledge G, Heward WL, Lo Y. 2006. Effects of response cards on disruptive behavior and academic responding during math lessons by fourth-grade urban students. *J Positive Behavior Interventions* **8**: 88–99.
18. Ma HH. 2006. An alternative method for quantitative synthesis of single-subject design: Percentage of data points exceeding the median. *Behav Modif* **30**: 598–617.
19. Mathai A, Provost S. 1992. *Quadratic Forms in Random Variables: Theory and Applications*. M. Dekker, New York.
20. Parker RI, Hagan-Burke S, Vannest K. 2007. Percentage of all non-overlapping data (PAND): An alternative to PND. *J Special Education* **40**: 194–204.
21. Shadish WR, Hedges LV, Pustejovsky JE. 2011. *An Application of a New d-Estimator for Single-Case Designs*. Paper presented at the Annual Conference of the Society for Research Synthesis Methodology, Ottawa, Canada.
22. Shadish WR, Rindskopf DM, Hedges LV. 2008. The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention* **2**: 188–196.
23. Shadish WR, Sullivan KJ. 2011. Characteristics of single-case designs used to assess intervention effects in 2008. *Behavioral Research Methods* **43**: 971–980.
24. Valentine JC, Cooper H. 2003. *Effect Size Substantive Interpretation Guidelines: Issues in the Interpretation of Effect Sizes*. Washington, DC: What Works Clearinghouse.
25. Wilkinson L. 1999. Statistical methods in psychology journals: Guidelines and explanations. *Am Psychol* **54**: 594–604.