Journal of School Psychology 63 (2017) 13–34

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Additional comparisons of randomization-test procedures for single-case multiple-baseline designs: Alternative effect types

Joel R. Levin [a,*,1], John M. Ferron [b,1], Boris S. Gafurov [c]

[a] University of Arizona, United States
[b] University of South Florida, United States
[c] George Mason University, United States

ARTICLE INFO

Article history:
Received 7 May 2016
Received in revised form 18 November 2016
Accepted 15 February 2017
Available online xxxx

Keywords:
Single-case intervention research
Multiple-baseline design
Randomization statistical tests
Alternative effect types

ABSTRACT

A number of randomization statistical procedures have been developed to analyze the results from single-case multiple-baseline intervention investigations. In a previous simulation study, comparisons of the various procedures revealed distinct differences among them in their ability to detect immediate abrupt intervention effects of moderate size, with some procedures (typically those with randomized intervention start points) exhibiting power that was both respectable and superior to other procedures (typically those with single fixed intervention start points). In Investigation 1 of the present follow-up simulation study, we found that when the same randomization-test procedures were applied to either delayed abrupt or immediate gradual intervention effects: (1) the powers of all of the procedures were severely diminished; and (2) in contrast to the previous study's results, the single fixed intervention start-point procedures generally outperformed those with randomized intervention start points. In Investigation 2 we additionally demonstrated that if researchers are able to successfully anticipate the specific alternative effect types, it is possible for them to formulate adjusted versions of the original randomization-test procedures that can recapture substantial proportions of the lost powers.

1. Introduction

Over the past several years single-case research methodology and associated data-analysis procedures have elevated their scientific "credibility" (Levin, 1994) among educational and psychological intervention researchers (see, for example, Kratochwill et al., 2013; and Kratochwill and Levin, 2014). The once-common single-case two- and three-phase AB and ABA designs, respectively, are now considered to be lacking scientific validity (Kratochwill et al., 2013) and so designs such as the ABAB "reversal" design, the alternate treatments design, and the multiple-baseline design have been advocated in their stead. Of these, many single-case intervention researchers (including the present authors) believe that the systematically staggered multiple-baseline design possesses the strongest internal-validity characteristics of all commonly adopted single-case designs in terms of its ability to document causal relationships between interventions and outcomes (see, for example, Horner and Odom, 2014; and Levin, 1992).

* Corresponding author.
E-mail address: jrlevin@u.arizona.edu (J.R. Levin).
Action Editor: Stephen Kilgus
[1] The first two authors contributed equally to this study. We are grateful to the reviewers of earlier versions of this article for their helpful revision suggestions.

http://dx.doi.org/10.1016/j.jsp.2017.02.003
0022-4405/

Yet, despite the high methodological marks accorded to single-case designs such as the multiple baseline, even greater scientific credibility can be attained through an interventionist's implementation of various forms of design randomization and data-analysis randomization to enhance the research's internal validity and statistical conclusion validity, respectively (Kratochwill and Levin, 2010). To emphasize the enhanced scientific credibility that accrues to single-case designs through researcher-managed *randomization* and *control*, and consistent with Shadish et al.'s (2002) orientation, we have started referring to designs that possess these two methodological components as "experimental" single-case intervention research designs (see also de Jong et al., 2008). It is implicit that all of the single-case intervention designs discussed in this article encompass randomization and control as two defining "experimental research" requisites.

The general topic of concern in the present simulation study is on statistical tests applied to data from multiple-baseline intervention studies, with a specific purpose of extending the results of a recently reported study on the statistical properties of a number of multiple-baseline randomization tests (Levin et al., 2016). The statistical properties in question are Type I error and power, and the randomization tests are ones that have appeared in the single-case literature throughout the past nearly half century. Single-case randomization tests have been gaining traction among educational intervention researchers, and notably, among researchers who have focused on academic and behavioral concerns (e.g., Ainsworth et al., 2016; Bice-Urbach and Kratochwill, 2016; Bardon et al., 2008; Brewer and White, 1994; Hwang et al., 2016; Lojkovic, 2014; Markham et al., 2011; Regan et al., 2005). The results of the present simulation study should be of direct relevance to single-case intervention researchers with those concerns.

1.1. Different types of single-case design-and-analysis randomization

Single-case intervention designs are interrupted time-series designs (e.g., Glass et al., 1975; McCleary, McDowall, & Bartos, in press) with at least two phases, a baseline or control phase (A) and an intervention or experimental phase (B), with each phase generally containing multiple outcome observations (Horner and Odom, 2014). In such designs, cases consist of either individual participants or clustered aggregates such as dyads, small groups, classrooms, communities, etc.; and the successive observations produced by each case typically are not independent, with the degree of nonindependence reflected by the magnitude of the autocorrelation coefficient. Because of the autocorrelated nature of the data, standard statistical procedures for assessing between-phase changes (for example, through parametric t or F tests) (1) do not satisfy the procedures' requisite assumptions, (2) will generally lead to unwarranted statistical conclusions, and therefore (3) should not be applied (see, for example, Ferron and Levin, 2014). Methodologically stronger single-case intervention designs are produced when they are replicated across cases—and consequently, commonly applied designs (e.g., AB, reversal, alternating treatment, multiple-baseline) must include a prescribed multiple-case replication component before they are endorsed by single-case intervention research "standards" committees (e.g., Kratochwill et al., 2013).

As we have indicated previously (Levin et al., 2014b), there are currently four different randomization variations that are being incorporated into single-case designs and statistical analyses: within-case intervention/phase-order randomization, between-case intervention randomization, case randomization, and intervention start-point randomization. Within-case intervention/phase-order randomization can be applied in ABAB…AB, alternating treatment, and simultaneous treatment designs (Levin et al., 2012); and between-case intervention randomization is manifested in situations where there is random assignment of one or more cases to one particular intervention (X) and another case or cases to a different intervention (Y), as in Levin and Wampold's (1999) "randomized pairs" AB design.

With the present study's focus on multiple-baseline designs, the third and fourth randomization types (case randomization and intervention start-point randomization) comprise the relevant types under consideration here. With case randomization, the N cases are randomly assigned to the N staggered positions, or "tiers" (e.g., Barton and Reichow, 2012), of the multiple-baseline design. With intervention start-point randomization, the within-position observation associated with the first session of the intervention phase is randomly selected from a pre-established number of potential intervention start points (k) that are "acceptable" to the researcher – an innovative process initially proposed by Edgington (1975) that has beneficial randomization-test consequences (Ferron and Levin, 2014). Thus, for example, in a four-case, 25-observation (O) multiple-baseline study, the researcher could specify that a minimum of 5 baseline (A phase) and 5 intervention (B phase) observations are required for each case, with two possible design variations represented by those specifications as follows: (*a*) the case within each multiple-baseline position receives a randomly selected intervention start point somewhere between O_6 and O_{21} inclusive (Marascuilo and Busk, 1988), resulting in a total of 16 potential intervention start points for each case;[2] or (*b*) the cases that are randomly assigned to Positions 1–4 are randomly assigned intervention start points falling within the intervals O_6–O_9, O_{10}–O_{13}, O_{14}–O_{17}, and O_{18}–O_{21}, respectively (Koehler and Levin, 1998).

1.2. Overview of five multiple-baseline randomization tests

Five different single-case multiple-baseline randomization-test design-and-analysis procedures (two of which included two different variations) were examined in the present simulation study, with three of the procedures involving intervention start-point randomization: Levin et al.'s (2016) restricted Marascuilo and Busk (1988) procedure, Koehler and Levin's (1998) "regulated

[2] The question of whether the N cases' intervention start points must be staggered is considered in a subsequent section.